Facultad de Informática

Trabajo de Fin de Máster en Ingeniería Informática para la Industria

Máster en Investigación en Informática, Facultad de Informática

Universidad Complutense de Madrid

Master's Thesis

ANALYSIS OF MEG SYNCHRONIZATION SIGNALS: APPLICATION OF SUPPORT VECTOR MACHINES AND CONFORMAL PREDICTION FOR CLASSIFICATION OF MILD COGNITIVE IMPAIRMENT

Thesis Advisor: Prof. Matilde Santos Peñas

Author: Juan García-Prieto Cuesta

Madrid, September 2011

El/la abajo firmante, matriculado/a en el Máster en Investigación en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo Fin de Máster: "ANALYSIS OF MEG SYNCHRONIZATION SIGNALS: APPLICATION OF SUPPORT VECTOR MACHINES AND CONFORMAL PREDICTION FOR CLASSIFICATION OF MILD COGNITIVE IMPAIRMENT", realizado durante el curso académico 2011-2012 bajo la dirección de Matilde Santos en el Departamento de Departamento de Arquitectura de Computadores y Automática, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Juan García-Prieto Cuesta

To Ma, to Me and to Ra.

Agradecimientos

Durante el último año he dado un vuelco al conocimiento que tengo del mundo en el que vivo. En particular, de la relación entre la inteligencia artificial y la natural. Aun puedo recordar la primera vez que mi profesora describía en pocas palabras y mediante un lenguaje cercano en qué consiste la inteligencia artificial. El hecho de que este tipo de estructuras se comporten como lo hacen, ha supuesto para mí un verdadero descubrimiento.

Por ello y por mucho más aprovecho esta oportunidad para agradecer a Dña. Matilde Santos por hacer realidad este proyecto de fin de Máster.

Así mismo me gustaría agradecer a D. Fernando Maestú por sus sabios consejos, por su paciencia y por haberme abierto la puerta a la neurociencia.

A D. Ricardo Bajo, sin su colaboración y su ayuda este trabajo nunca hubiera existido.

Me gustaría agradecer al Centro de Tecnología Biomédica de la Universidad Politécnica de Madrid, personalizado en D. Francisco del Pozo y al departamento de Neurociencia Cognitiva y Computacional por su inestimable ayuda para la consecución de todos los objetivos propuestos para este trabajo, así como la cesión de los datos estudiados con completa generosidad y entendimiento.

Septiembre 2011

Executive Summary

Keywords: Alzheimer Disease, Mild Cognitive Impairment, Magnetoencephalography, Functional Connectivity, Likelihood Synchronization, Support Vector Machine, Conformal Prediction, Recursive Feature Elimination

Alzheimer Disease and Mild Cognitive Impairment are an issue of serious global concern. Scientific progress has brought great benefits to the modern society; meanwhile the rapid increase in life expectancy has raised difference neurophysiology issues. The idea that appears in the horizon is clearly the possibility of an earlier diagnose of Alzheimer illness. The importance of early diagnose becomes critical due to the fact that neurological systems do not recover from degeneration. Therefore, an early detection and subsequent treatment could at least slow the cognitive disrepairment.

There have been very satisfactory results while applying Likelihood Synchronization algorithm due to its particular capabilities towards time-space-synchronization capabilities. Since then, once introduced the processing algorithm to MEG data, synchronization gives an index which provides a nonlinear characterization of functional connectivity.

This work represents a first approach from an Artificial Intelligence perspective by means of applying Machine Learning Support Vector Machine techniques in a twofold manner: mainly as an effort to develop a classifier based on MEG recordings and to try to develop a standardized procedure for dealing with MEG Synchronization recordings; and secondly a possible approach to extract information of the brain functioning inferred from the previously trained classifier. Trained SVM were used in order to apply Enhanced Recursive Feature Elimination techniques and weight the influence of single Synchronization Likelihood links on each classifier decision. Conformal prediction will grow a layer of credibility and reliability to prediction results.

Final results show a top 86% correct ratio by leave one out cross validation training scheme, with linear, polynomial and radial basis kernels.

Spanish Summary. Resumen

Palabras clave: Enfermedad de Alzheimer, Deterioro Cognitivo Leve, Magnetoencefalografía, Conectividad Funcional, Sincronización Likelihood, Máquinas de Vectores de Soporte, Predicción Conformal, Eliminación Recursiva de Características

La Enfermedad de Alzheimer adquiere una importancia cada vez mayor para el ser humano. Su diagnostico precoz como por ejemplo en la fase conocida como síndrome de Deterioro Cognitivo Leve, es útil para el tratamiento de dicha enfermedad.

La magnetoencefalografía es una técnica relativamente moderna, que adquiere relevancia por su idoneidad para el estudio de la dinámica de las redes cerebral desde una perspectiva de Conectividad Funcional. Mediante el tratamiento de datos magnetoencefalográficos pertenecientes a una prueba neuropsicológica y su estudio desde una perspectiva de Conectividad Funcional a través del algoritmo de Sincronización Likelihood, el presente trabajo pretende abordar el estudio de dichas señales para aplicar Máquinas de Vectores de Soporte y desarrollar un método de clasificación eficaz de pacientes con el síndrome de Deterioro Cognitivo Leve.

Diferentes estrategias de pre-procesamiento serán estudiadas valorando sus resultados de manera paralela para intentar desarrollar un procedimiento estandarizado que sirva tanto para el método de clasificación utilizado en este trabajo como para posteriores aplicaciones. Diferentes funciones *kernel* serán valoradas para el conjunto de test y sus resultados y eficiencia de ejecución sometida a estudio comparativo.

Así mismo, mediante el uso de la teoría de Predicción Conformal, la clasificación mediante Máquinas de Vectores de Soporte incorporará una medida cuantitativa de la fiabilidad y la credibilidad de cada predicción.

Los resultados arrojan la cifra de 86 % de predicciones positivas para el conjunto de datos estudiado, mediante una estrategia de entrenamiento que minimiza el riesgo de sobre-entrenamiento denominada *leave one out cross validation*.

Por último, los resultados obtenidos serán evaluados mediante una técnica de eliminación recursiva de características, para permitir valorar desde una perspectiva neuropsicológica los diferentes links de sincronización cerebral más relevantes para la función de clasificación. Permitiendo de este modo, un posterior contraste de hipótesis sobre los efectos de la Enfermedad de Alzheimer y sus efectos en diferentes procesos cognitivos.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABREVIATIONS

AD	Alzheimer's Disease
AI	Artificial Intelligence
CP	Conformal Prediction
EEG	Electroencephalography
EnRFE	Enhanced Recursive Feature Elimination
fMRI	functional Magnetic Resonance Imaging
HA	Healthy Ageing
HDLS	High-dimension Low-sample size
LOOCV	Leave One Out Cross Validation
MCI	Mild Cognitive Impairment
MEG	Magnetoencephalography
MRI	Magnetic Resonance Imaging
MSR	Magnetic Shielded Room
PET	Positron Emission Tomography
RFE	Recursive Feature Elimination
SL	Synchronization Likelihood
SQUID	Super Quantum Interference Device
SVM	Support Vector Machines

INTRODUCTION

One of the most ambitious and challenging scientific problems of all time has been to understand the details of brain structure. There have been many advances during past decades, in the direction of understanding where different functions of the brain are mapped. With mayor improvements concerning neuroimaging techniques, different parts of our brain have been defined. It has been demonstrated by many empirical observations, how each different brain region has a specific function. (T. M. Jessell 2000).

The next step in the study of the brain has become the understanding of how different parts of the brain interact with each other. Brain processes are dynamic and imply different brain parts interaction for cognitive tasks to be accomplished. Therefore, there has been a great effort focused on the integration of brain activity in a meaningful model. And the temporal dynamics of neurophysiological signals make the analysis of brain activity a true challenge.

One main approach facing this challenge is the concept of Functional Connectivity, which refers to the statistical interdependencies between physiological time series recorded in various brain areas simultaneously and is probably, an essential tool for the study of brain functioning, being its deviation from healthy reference an indication of lesion.

Yet, alone this might seem important enough, its approach through the study of brain diseases may well be worth the effort of trying to develop new applications of advances in different areas of machine learning theory. Integrating different classification techniques enrich neuroscience studies and particularly Alzheimer Disease (AD) or Mild Cognitive Impairment (MCI) as a previous stage. Artificial Intelligence (AI) and supervised learning can recognize patterns beyond data and estimate qualities in it, inferencing from a training data set.

This work is a binary classification exercise. Its main concern is to develop a set of tools for classification and early diagnose of AD, particularly on its early stage, and to add a certain level of confidence to the classification result. To accomplish this, Magnetoencephalography (MEG) data has been used, through a Synchronization Likelihood (SL) analysis which underlines functional connectivity between brain areas. Later all these concepts will be attended.

About the study of brain diseases

AD is the most common dementia in the elderly and is estimated to affect 35.6 million people worldwide. AD is believed to have a prodromal stage lasting ten or more years. The incidence and prevalence of AD begins to rise as individuals reach the age of 65 such that by the time they are in their 80s and 90s, the risk of clinical dementia is nearly 50%. However, due to the fact that the risk of the clinical syndrome, Alzheimer's dementia, is greatest in the later years of life, pathological processes begin ten to twenty years before clinical onset. This means that treatment strategies aimed at disease modification will be most effective if they can take place during the period when the pathological changes are occurring, but have not yet exhibited themselves as clinical signs and symptoms.

Diagnostic criteria for AD has been well codified since the early 1980s, there has been a recent upsurge in interest in studying individuals who are in the transitional stage between normal cognition and full-blown dementia. This syndrome, referred to as Mild Cognitive Impairment (Flicker et al., 1991;Petersen,2004), has been the focus of intense study and there are many who believe that in the absence of other medical comorbidities individuals with MCI, in fact have clinical AD in its earliest stages. MCI is defined as a clinical condition characterized by memory impairment and deterioration of additional cognitive domains, which do not interfere with daily living activities; (i.e. when cognitive impairment is not severe enough to constitute dementia). Early identification of patients at risk for the development of dementia might be crucial for proving them cognitive or pharmacological interventions with the aim of slowing the progression of cognitive deficits and to retard the onset of disability (Braak et al., 1991).

It is believed that the majority of MCI patients that convert to dementia do so within 10 years. Early AD diagnose importance, lies on two main reasons: i) Once EA has been developed and it's symptomathology is clear, brain damage is deep and severe. Therefore it is then unlikely a successful treatment not yet a recovery. ii) Underlying pharmacological development of AD treatments, improvement in AD markers can open a window to effectively explore relevant responses to treatments.

Machine Learning

Pattern Recognition looks like an immensely broad subject with countless applications. One of the things I feel most attracted to, regards how it can be used as a method both for eventually classifying samples among different groups, and also (perhaps) as a way to understand a little bit more about each group. Classification is, at base, the best task of recovering the model that generated the patterns.

A simple search under "machine learning brain" throws as much as 5166 *pubmed* bibliographic results, though only 59 of them consider MEG data analysis, and only 4 focus on Alzheimer's disease. However, there is rapidly accumulating evidence that the application of machine learning classification to neuroimaging measurements may be valuable for the development of diagnostic and prognostic prediction tools (Nouretdinov, I., 2010).

The purpose within this work is to reliable classification between MCI and Control subjects will be attempted. And work towards an as-systematized-as-possible method for analyzing SL - MEG data from a memory task.

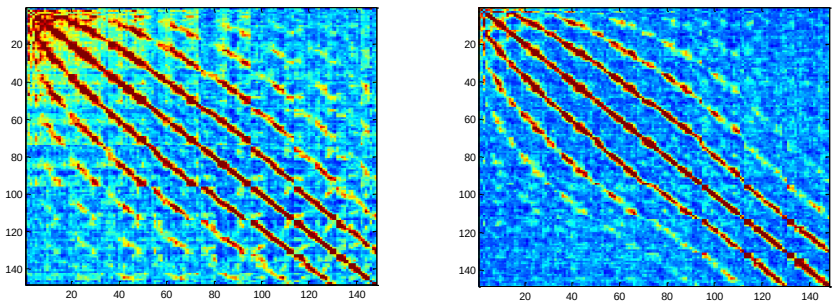


Figure 1. Two example of Synchronization Likelihood symmetric output images. Each row or column represent one in 148 possible channels, and SL value is given between 1 and 0. Left, corresponds to a typical MCI subject where right image matches a typical Control subject output.

Data set will consist of 22 subjects belonging to MCI group and 19 to the control group, and as each SL instance consists of 10878 dimensions[1]. Data set is a particularly good example of what a high-dimension low-sample size (HDLS) data set is. Therefore, given 10878 variables for each sample, there must truly be differences between samples for an effective classification to take place. Information from single subject will be analyzed with feature extraction algorithms whose purpose will be to reduce data dimensionality while maintaining essential properties in the data.

Several different methods and procedures are attended through this work with a main objective of developing an efficient diagnose tool. Particularly, Support Vector Machines (SVM) will be used extensively. The ability of SVM to avoid Structural Risk Minimization[2] is a main idea in their theory and method. However perfect classification is simply impossible and therefore, there is an overall cost associated with each decision. The true task of this work is to try to minimize such a cost or at least try to know the expectable confidence on each prediction. Conformal Prediction (CP) will be used in order to achieve this sense of relyness.

Why MEG

Magnetoencephalography is a reasonable young technique, particularly suitable for studying brain dynamic behaviour. The high temporal resolution of the MEG technique allows measuring the dynamics of the oscillatory activity and as a consequence establishing the functional interaction between brain areas at specific frequency bands. Therefore, MEG provides a four dimensional view of brain function (space-time-frequency-connectivity) which open a better description of the consequences of neurological diseases on the functional networks which support cognitive functions. MEG may have potential as a biomarker of AD and we must evaluate the relative merits of the methodology in neurodegenerative disease. For this reason the use of MEG technique is proposed as a very valuable technique to study the functional networks due to AD or its preclinical states.

Designing a classifier has the central aim of suggesting actions when presented a “novel” pattern, and this way getting over the issue of generalization. During the first chapter a brief explanation of SVM theory will be exposed. After the meaning of support vector is understood, CP will be added to the method in order to provide a further layer of confidence in the results. CP uses past experience to determine precise levels of confidence in new predictions. In other words, to try and answer questions as: how good a prediction is or what confidence can you expect from a prediction.

Brief Chapter Description

After this introduction on the state of the art in Functional Connectivity, particularly, when research focuses on AD and MCI, a short review of the fundamental theory that underlines SVM and CP is developed in chapter 2.

[1] This will be later explained.
[2] Minimizing the generalization error or risk.

As a multidisciplinary effort, this work requires to know the background of the data set, in order to be able to satisfactory read into the information contained inside it. MCI – Control classification is strongly determined by neuroscientific context. Because of the particularities that related to working with a MEG machine, to deal with subjects, out layers, and neuroscientific evaluations, this work needs a deep understanding of contextual meaning of a considerable amount of concepts. All these are explained in chapter 3. After which, what is truly considered *raw data* is completely and correctly understood.

After this background is assumed, different techniques will be proposed to either reduce data dimensionality or try to improve the way information can be extracted from it. It will become useful to achieve a representation of synchronization data, as a way to start performing feature extraction. This effort will end up with a complete set of parallel approaches, as different pre-processing techniques as a way to validate the optimal itinerary towards a satisfactory classification.

Chapter 4 will show classification results for each of the different pre-processing techniques proposed in the previous chapter. Along this work multiple classifiers will be tested while trying to perform an insightful comparative analysis of each method. Eventually a collection of highly efficient methods will be put aside to try and perform CP based on them.

Because of the data set to be considered low-sample size related to the number of samples, dimensionality reduction based on SVM methods will we tested.

Finally in chapter 5 will explain different conclusive results. As a multidomain effort this work will try to underline an eventual systematic method for analysing MEG data, and also while this is done, advance eventual following developments to be performed.

STATE OF THE ART

Support Vector Machines

Though centred on the neuroscience problem related MCI diagnose, during the next pages a brief explanation of what Support Vector Machines are and what they can do about the pattern recognition problem will be attended.

On a given learning task like this one, with a given finite amount of training data, the best generalization performance will be achieved if the right balance is struck between the accuracy attained on that particular training set, and the "capacity" of the machine, that is, the ability of the machine to learn any training set without error. Too much capacity ruins learning because of lack of generalization; but too little capacity simplifies too much, avoiding "retention" of details and therefore ruing learning because of too much generalization.

Given *l* observations, consisting of pairs of observations and the associated label:

$$l\ observations\ with\ x_i \in \mathcal{R}^n, i = 1, \dots, l$$

$$associated\ with\ a\ label\ (truth)\ given\ by\ a\ trusted\ source^{3}$$

$$y_i \in \{-1,1\}$$

Data is assumed to be "iid": independent and identically distributed while selected from a universe from a trusted source. [4] There exists a *P(x,y)* probability distribution which is more general than associating a fixed *y* with every *x*.

We will suppose we have a machine that maps $x_i \rightarrow y_i$ with some α adjustable parameters: $x \rightarrow f(x, \alpha) = y$. To train the machine literally means to decide which α parameters are more suitable for each problem.

SVM have very good generalization performance without the problem of *course of dimensionality*[5], which affects other families of machine learning.

[3] By trusted source it is understood that prior to computational work, or even during an earlier stage in the process, trusted neuropsychological assessment will evaluate the belonging group of each one of the subjects taking part in the study. A brief description of the process is given in Chapter 3.

[4] It is widely accepted from a statistical point of view, for the subjects randomly selected during clinical evaluations to be independent and identically distributed.

[5] Other methods of machine learning suffer from a wide variety of difficulties such as "the course of dimensionality", which refers to the way complexity of a learning methods shows, while complexity is added to the problem (i.e. adding more dimensions).

Given N training samples $\{x_i, y_i\}, i = 1, \dots, N$ *and* $y_i \in \{-1,1\}$

where $x_i \in \mathcal{R}^n$ *and* y_i *is the class label.*

Firstly, let's consider the linear separable case where the two groups are separable with a straight line, which will define the separating hyperplane. The classification problem will be formulized by a function:

$$f(x, \alpha) = (w_\alpha \cdot x) + b \tag{1}$$

Among the possible hyperplanes SVM will focus in finding the optimal separating hyperplane, and this is determined by the vectors on the margin, called support vectors.

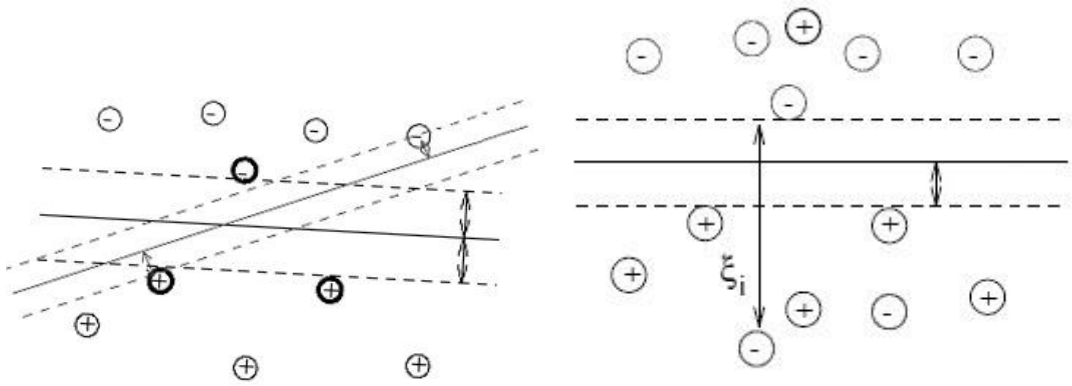


Figure 2 SVM training examples. Left: Linearly separable case. Were the maximum margin hyperplane is plotted while SVM learning process. Right: Non-separable data sets obey to introduce the concept of *soft-margin*.

With,

$$y_i[(w \cdot x_i) + b] \geq 1 \tag{2}$$

When data is not separable, training may still take place if a new weight ξ_i is introduced, to allow for some error. (The higher the weight, the higher the penalty for allowing training errors.)

$$y_i[(w \cdot x_i) + b] \geq 1 - \xi_i \quad with \;\; \xi_i \geq 0 \tag{3}$$

For the non-linearly separable case, SVM first maps the data to another Hilbert space $\mathcal{H}$ (feature space) using a mapping function $\Phi: \mathcal{R}^n \rightarrow \mathcal{H}$ that satisfies Mercer's conditions.

In the feature space $\mathcal{H}$, we find an optimal hyperplane by maximizing the margin between training samples and bounding the number of training errors. The decision function can be driven by,

$$f(x) = \theta(W \cdot \phi(x) - b) = \theta(\sum y_i \alpha_i \, \phi(x) \cdot \phi(x_i) - b) = \theta(\sum y_i \alpha_i \, K(x_i, x) - b) \tag{4}$$

Where α_i defines if x_i is a Support Vector as it will be $\alpha_i = 0$ for every non-Support Vector.

Note the importance of the switch that the use of *K* function implies. Instead of calculating the dot product $\phi(x) \cdot \phi(x_i)$ of the pair of vector in the higher dimensional Hilbert feature space, a "trick" is used in order to avoid such task. As previous expression only depend on the dot

product of both vectors in the feature space, a Kernel function is used to obtain the same result. This means:

$$\phi(x_i) \cdot \phi(x_j) = K(x_i, x_j) \tag{5}$$

Training the SVM means finding $\alpha_i, i = 1, \cdots, N$ that maximize

$$L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \alpha_i \, \alpha_j y_i y_j K(x_i, x_j) \tag{6}$$

$$With \; 0 \leq \alpha_i \leq C \; in \; i = 1, \cdots, N$$

$$and \; \sum_{i=1}^{N} \alpha_i \, y_i = 0$$

Where C is a parameter that defines the penalty for training errors. And N is the number of support vectors.

From the last equation with constraints, it can be assumed that the optimal separating hyperplane is defined by the so-called support vectors. For these $\alpha_i \neq 0$. However values under an epsilon level are usually ignored (epsilon insensitive zone). This plays a major role in the accuracy of SVM classification. And even more importantly: the description of the optimal separating hyperplane does not explicitly depend on the dimensionality of the problem.

The solution of the optimization problem has the form:

$$\boldsymbol{w} = \sum \alpha_i y_i \boldsymbol{x_i} \; and \; b = y_i - \; \boldsymbol{w^T x_i} \; for \; any \; \boldsymbol{x_i} \text{ such that } \alpha_i \neq 0 \tag{7}$$

The outstanding generalization ability for the SVM comes mainly from two reasons. First, the optimal separating hyperplane maximizes margin between groups. And secondly, the small number of Support Vectors, relative to the sample number can enhance generalization capacity.

There are no guarantees of generalization accuracy. Even though SVM theory is based on the idea of Structural Risk Minimization, there no notion of *confidence* in a determined prediction. This is exactly the gap covered with CP, described as follows.

Conformal Prediction

Conformal Prediction relies on confidence intervals from classical statistics and are well theoretically founded (Nouretdinov, I., 2010).

CP creates a layer of relyness (confidence) on top of other methods of prediction by using past experience to produce a set of labels (Shafer and Vovk, 2008). CP can produce precise levels of confidence in new predictions, given an error probability of $\boldsymbol{\varepsilon}$ together with a method that makes a prediction $\hat{y}$ of a label$y$, it produces a set of labels that contain $\hat{y}$ with a $\boldsymbol{\varepsilon}$ probability.

CP is a type of machine learning methods defined in a transductive and on-line framework. By transduction it is meant that there is no need for a general role induction and a deduction for a new instance in order to perform a prediction; CP directly *transducts* from a training set a determined probability of a prediction. By online it is meant to show that the particular way in which CP works, makes it easy to implement a framework in which each new prediction is bases on all previous samples instead of using a rule constructed from a fixed batch of trails.

As the explanation of SVM started with, given l observations, consisting of pairs of observations and the associated label

$$l\ observations\ with\ x_i \in \mathcal{R}^n, i = 1, \dots, l$$

$$associated\ with\ a\ label\ (truth)\ given\ by\ a\ trusted\ source$$

$$y_i \in \{-1,1\}$$

CP will predict the label fory_{l+1}. The only assumption is to consider the data *iid*[6] , and to carry out the prediction, conformal predictors try every possible label $y_i \in \{-1,1\}$ as a candidate for the label of the new sample. How well each possible label conforms to the randomness assumption will determine which label is predicted. Ideally, only one case will lead to sequences that are not random, and a measure of reliability of the prediction (confidence and credibility) will be accomplished.

One of the main benefits of CP is that there is really no difference between learning and prediction, as all objects are treated simultaneously and CP learns and predicts at the same time. Confidence measures are obtained at each classification without relying on values obtained with a fixed subset of samples, and ambiguities in the classification process can be detected. For CP to classify it is necessary to measure how different the new sample is from old (label-known) samples. To do this, a non-conformity measure is needed, and a non-conformity score will be determined. In our case, the non-conformity measure will constitute the α_i values of the SVM training.

Given a non-conformity measure it is possible to compute a non-conformity score for each possible label, and after that a p-value will compare all α_i to determine how different or similar the new sample is from the initial set. Compares the non-conformity score of the new sample with all other non-conformity scores. For a wrong prediction we expect the non-conformity score to be a little bit higher, than if correct. In such case, a low p-value will be obtained.

For the true label of x the p-value function satisfies the following property for all probability distributions P and for any significance level ε:

$$P(p(y_{n+1}) \leq \varepsilon) \leq \varepsilon \tag{8}$$

The property describes that when the given training set contains iid instances, the probability of the p-value of the training set to be less than or equal ε is less than or equal ε. Consequently we may output a set of possible predictions (i.e. a predictive region= which contains all the

[6] iid: independent and identically distributed. This will refer to the fact that the probability distribution for each label is unknown, but as long as it is iid, it will be bounded under certain known limits.

prediction with p-values greater than the significance level ε. Moreover, we always include the highest prediction in order to ensure that the predictive region will contain at least one prediction.

Because of the property, the probability of each set S not containing the correct prediction will be less than or equal to ε. As a result, the error of the predictive regions will be bounded to ε and thus we can say that we have a 1-eps confidence in our predictions.

Alternatively, the CP may output a single prediction which is the prediction with the highest p-value complemented with a confidence measure which is 1-second p-value, and a credibility value which is the p-value of the prediction. The confidence measure shows how likely the output classification is of being correct, compared to all other possible classes.

The credibility value gives an indication of how suitable the training set is for classifying the current instance. Label prediction will be the largest p-value. The credibility will be the largest p-value and the confidence in the prediction will be 1-2nd p-value.

ANALYSIS OF THE PROBLEM

In order to understand nuances which might be hidden inside MEG data, and before going into the pre-processing procedure, a brief description of the most important stages involving SL analysis of MEG data will be exposed along this chapter. For a brief description of the setup regarding MEG apparatus go to Appendix B.

To begin with it is essential to address the definition and understanding of the processing unit that will apply in this work. "Epoch" is in standard neuropsychology, a way to refer a critical time gap just after a stimulus is presented to a subject. However, "epoch" is a very common way of referring a "learning state" in standard artificial intelligence theory. Following references to either concept will be carefully explained to avoid any confusion.

Epoch: **An epoch in neuroscience is a section of the incoming continuous data, defined by the occurrence of an event and the time limits with respect to that event. Traditionally machine learning methods refer to 'epoch' as each stage of the training process; however the first definition will attend here.**

This chapter will explain where the raw data comes from. The hardware being used, and how SL applied over to certain epochs, might expose hidden and plain features of brain's Functional Connectivity, are all concepts which will explain mainly the *what* yet, not forgetting the *why*, of data in question.

Meg setup

MEG recordings were performed with a whole-head neuromagnetometer consisting of 148 magnetometers coils. The instrument is housed in a magnetically shielded room (MSR) designed to reduce the environmental magnetic noise that might interfere with biological signals.

Figure 3 A MEG registry taking place, while at up-straight position.

All sensors included inside a DEWAR container with liquid Helium (4.2K). Signals come out SQUID sensors and go through a Signal Acquisition System were they are conveniently processed and digitalized. The neuromagnetic field they sense is associated with intracellular currents which occur due to postsynaptic voltages. These are usually modelled as a current dipole. The power of an equivalent current dipole is typically 10^{-14} Am, which produces, applying Amperes Law, a —typically— magnetic field of 10^{-18}T, in a distance of 3-5 cm from the neuron. Moreover, it has been estimated that in order to obtain a reasonably good SNR, each SQUID sensor in the Magnetoencephalograph will register magnetic fields generated by groups of 10^5 neurons which are activated together during brain normal activity. The magnetic field will be sufficiently weak as to avoid any detection for distances further than 3-4 cm.

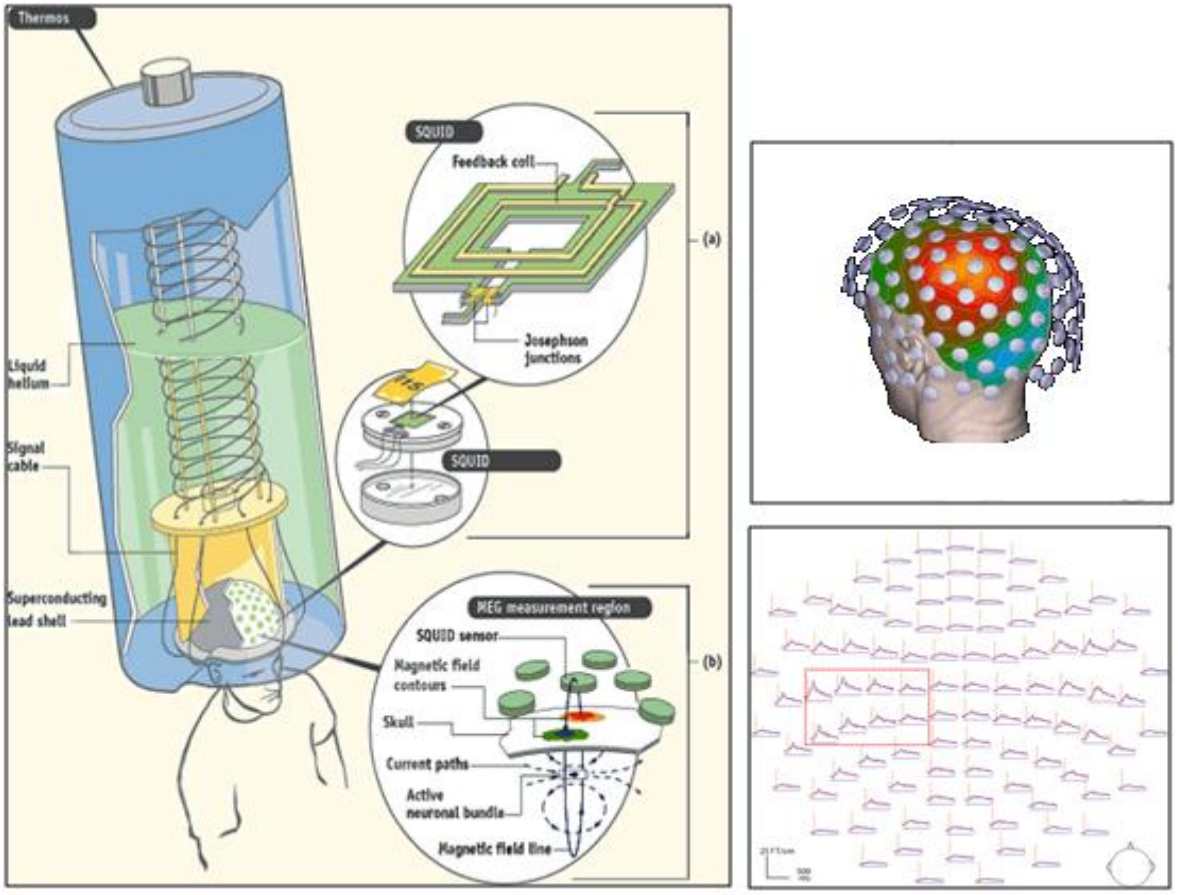


Figure 4. . Left and Top Right: Schematic descriptions of MEG setup. Down Right: Representation of sampled epochs during a normal MEG registers, in each one of the 148 sensors

The principal challenge of the bio-magnetic science is to be able to effectively measure magnetic fields as weak as 1fT up to 100pT, where the noise can be measured bigger than 10μT. Hence there is need for sophisticated techniques of noise cancellation.[7]

[7]There has been several publications regarding technical aspects of bio-medical instrumentation. One extense review can be found in NATO-ASI book, edited by Weinstock (1996).

The setup build by the University Complutense integrated a MEG: Magnes 2500WH, which includes 148 SQUID sensors inside a MSR.

The signal was filtered online with a band pass filter between .1Hz and 50Hz, digitized at 254Hz sampling rate. These steps are necessary to minimize the amount of low frequency magnetic noise that is typically present in MEG recordings. Epochs will be then saved together after removing those during which an excessive movement or blink had occurred.

MEG registries can sample spontaneous brain activity as well as responses to certain stimulus like audio, visual and sensitive. Sometimes connecting different structural properties or functions, other times focusing in mapping singular behavioural spots. All this give psychiatrists and neuroscientists a valuable tool for the study of cognitive functions and several clinical procedures.

Stimuli and Task

A modified version of the Stenberg's letter-probe task (de Toledo-Morrel et al., 1991; Maestú et al., 2001) was used. A set of five letters was presented and the participants were asked to keep the letters in mind. After the presentation of the five-letter set, a series of single letters (1000ms in duration with a random ISI[8] between 2-3s) was presented one at a time, and the participants were asked to press a button with their right hand (with a sole finger movement in order to minimize transitory magnetic noise being recorded) when a member of the previous set was detected.

The list consisted of 250 letters in which half were targets (previously presented letters), and the remaining letters were distracters (different from the previously presented letters). All participants completed a training session before the actual test, which did not start until the participant demonstrated that he/she could remember the five-letter set.

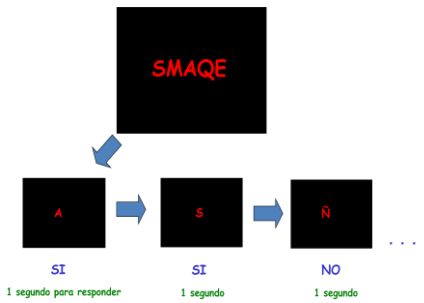


Figure 5 Example of the modified Steinberg stimulus

Letters were projected through a LCD video projector, situated outside of the magnetic shielded room, on to a series of in-room mirrors, the last of which was suspended approximately 1m above the participant's face. The letters subtended 1.8 and 3degrees of horizontal and vertical visual angle respectively.

[8] ISI: Inter-stimulus interval

Participants

Forty-one right handed, elderly participants recruited from the Geriatric Unit of the Hospital Universitario San Carlos (Madrid) participated in the study. Participants were divided into two groups based on a clinical neuropsychological profile. Twenty-two participants were diagnosed as MCI subjects and nineteen as health aging participants (HA), or simply control volunteers, without memory complaints.

MCI diagnose was established according to the criteria proposed by Petersen et al. (Petersen, 2004).[9] Patients and controls underwent a neuropsychological assessment, in order to establish their cognitive status in multiple cognitive functions, and according to their clinical and neuropsychological profile all participants in this group were considered multi-domain MCI patients.

MEG recordings

The MEG signal was recorded with a 254Hz sampling rate and a band pass filter between 0.5 to 50Hz was used. Using 148-channel whole head magnetometer, confined in a MSR. An environmental noise reduction algorithm using reference channels at a distance from the MEG sensors was applied to the data. Thereafter, single trail epochs[10] where visually inspected by an experienced investigator and epochs containing visible blinks, eye movements or muscular artefacts were excluded from further analysis. Artefact-free epochs from each channel were then classified into four different categories according to the subject's performance in the experiments: hits, false alarms, correct rejections and omissions. Only hits were considered for further analysis because we were interested in evaluating the functional connectivity patterns which support recognition success. 35 epochs were used to calculate SL values. To have an equal number of epochs across participants, 35 epochs were randomly chosen from each of the other participants.

MEG Functional connectivity: Synchronization Likelihood.

The Synchronization Likelihood method was developed by the group of Stam CJ (Stam et al, 2002), as an index which provides a nonlinear characterization of functional connectivity. However, it has been widely studied by other groups, having adapted this method to measured

---

[9] This criteria will constitute the later referred reliable third party in conformal prediction. MCI diagnosis will be established according to the criteria proposed by Petersen et al (Grundman et al, 2004; Petersen, 2004). Thus MCI patients fulfil the following criteria: 1)Cognitive complaint corroborated by informant; 2)Objective cognitive impairment, documented by delayed recall from the Logical Memory II subtest of the Wechsler Memory Scale Revised; 3) Performance on all measures of cognition must be > 1.5 standard deviation units away from expected value (i.e. no more than mildly impaired). Additionally, general cognitive function will be determined by clinician's judgement based on a structured interview with the patient and an informant.

[10] As noted previously "epoch" here, refers to new instances of the stimuli.

event related activity what we called Event-Related Synchronization Likelihood (Bajo et al., 2010).

In this method, as in any other form of synchronization, we will calculate the synchronization value between all pairs of channels. First of all, it is necessary to normalize (between 0 and 1) all the sampled data.

As an example we will select only a couple of channels. Let's consider channel number 12 and channel number 100. We then calculate the synchronization value between them in a period of 10 seconds.

We will divide each one of these channels into *pieces* vectors of length $\times\, l$;

$$X_i = (x_i, x_i + l, x_i + 2l, \cdots, x_i + (m-1)l) \tag{9}$$

The lag l determines the distance between the sample taken in the vector and the embedding dimension m is the length of each constructed vector.

Subsequently we choose a window with dimension W1 and W2. These two values are the window boundaries; the lower window boundary W1 is the Theiler correction[11] to remove autocorrelation effects and the upper window bound W2 is used to sharpen the time resolution of the synchronization measurement. Together they represent how close together in time the vector should be compared.

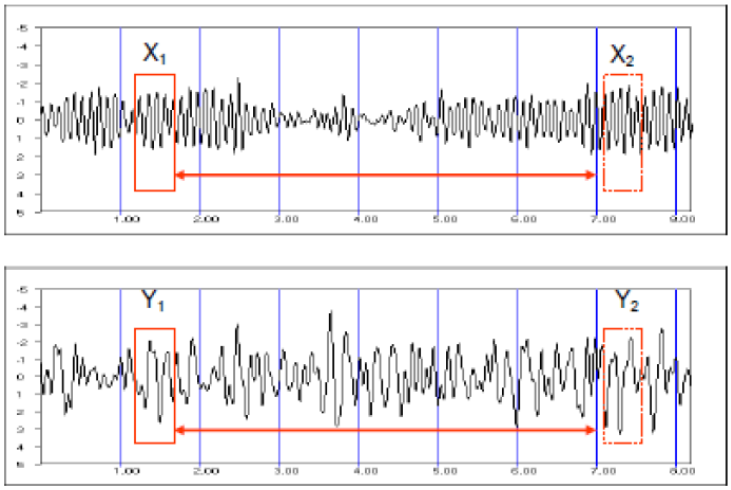


Figure 6 Example of how Synchronization Likelihood algorithm finds repeated correlated patterns among different channels.

[11]A correction for samples close to a reference point in order to reduce the influence of linear correlation on nonlinear measures.

We will calculate the probability of X2 to be close (Euclidean distance) to X1 at the same time Y2 is close to Y1.

Next, it is necessary to calculate the Euclidean distance between the window W1 and W2 and the rest of the pieces inside the mentioned vectors #12 and #100, as well as to check which one of these subtractions was less than an epsilon value:

$$\text{Channel \#1 } H_{a,b}^{1} = \theta(\varepsilon_{1a} - |X_a - X_b|) \tag{10}$$

$$\text{Channel \#100 } H_{a,b}^{100} = \theta(\varepsilon_{100a} - |X_a - X_b|) \tag{11}$$

$$\text{Where } W_1 < |a - b| < W_2$$

$$b \in (1, \cdots, N) with\ N \in \mathbb{Z}$$

$$\theta\ is\ the\ Heaviside\ Step\ function$$

$S_{1,100}^{a}$ is the variable we counted how many times the Euclidean distance is $H_{a,b}^{1} = H_{a,b}^{100} = 1$ simultaneously smaller in both channels than the epsilon value ($|X_a - X_b| \approx 0$) Thus we calculate the Synchronization Likelihood for "piece" a, between channel #12 and channel #100.

$$\left[S_{total}\right]_{channel\#1,channel\#100} = \frac{S_{1,100}^{a}}{2\cdot(W_2 - W_1)} \tag{12}$$

Finally the total value of the Synchronization Likelihood, between channels #12 and #100

$$\left[S_{total}\right]_{ch\#1,ch\#100} = \frac{1}{N}\sum_{i=1}^{N}[S_{total}]_{ch\#1,ch\#100}^{i} \tag{13}$$

Where N=Total number of points = (sample frequency) (10 seconds).

Obviously the process must be repeated for each pair of channels, therefore for 148 different channels the synchronization analysis sums a total of 21904 different SL values.[12]

Overall data view

The study relies on MEG registers of 19 subjects belonging to a control group and 22 subjects with MCI positively diagnosed. For each stimulus presented to the subject with subsequent positive hit, an epoch of 10s was defined. A SL analysis was performed, to each epoch, rejecting high noise epochs. Next figure shows a representation of the data set.

Next table shows an overview of the raw data considered in this classification exercise.

[12] Yet symmetrical properties will downgrade the number of different SL values to 10878. This is not relevant now, but will be important when classification gets started. Starting with a 148 by 148 analysis, we know that symmetric comparisons will have the same SL value, and we also know that each cannel with itself will have a SL value of '1'.

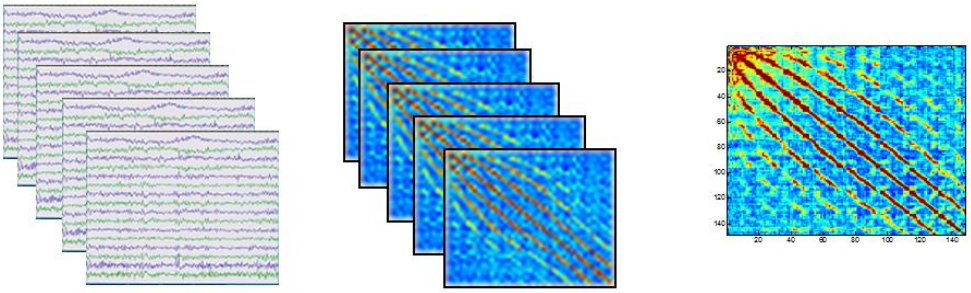


Figure 7 represents a typical result for a SL analysis. From left to right, 10 second epochs are analyzed with SL algorithm. Right: an example of a SL output, from a control subject. Note how neighbour channels have a higher statistical synchronization, which appear as semi-parallel stripes.

CONTROL GROUP		MCI GROUP	
subject	***number of epochs***	***subject***	***number of epochs***
1	35	**1**	35
2	35	**2**	35
3	30	**3**	35
4	35	**4**	35
5	35	**5**	35
6	35	**6**	35
7	35	**7**	30
8	35	**8**	35
9	35	**9**	35
10	35	**10**	35
11	35	**11**	35
12	35	**12**	35
13	35	**13**	29
14	35	**14**	35
15	35	**15**	35
16	35	**16**	35
17	35	**17**	35
18	35	**18**	35
19	35	**19**	35
		20	29
		21	35
		22	35

Table 1 Raw data considered for this work.

ALGORITHMS AND IMPLEMENTATIONS

Data analysis and pre-processing

Label Group	# subjects	# of total samples
Control	19	660 epochs
MCI	22	753 epochs

Table 2. A condensed resume of all data considered

Data set is not symmetric inter group. Therefore a deep analysis of the data will be performed in order to try to find out a good de-noising technique, which can reduce data dimensionality in order to improve later training efficiency.

Previous work related to this data (Bajo et al 2010), has used a mean reduction in order to compress all 35 epochs[13] in one single instance but trying to retain as much features as possible from all the epochs belonging to that subject. Proceedings in this work will follow that trace, but still, an effort will be made in order to try to improve that method. Research on various different methods to compress and de-noise all instances of a single subject into one single epoch will follow.

But first, if the mean is going to be used, it should be useful to graph randomly chosen SL values behaviour during all 35 epochs. Next page will show how it no mayor variations usually occur along epochs, in terms of making a mean reduction an undescriptive measure. Therefore, mean can be considered a good de-noising method, however, as shown, some SL values have rather high variation.

It is possible to map this different variation rate, by mapping a 148 by 148 matrix with normalized standard deviation during epochs for every single matrix component. Inside Appendix A this analysis is performed to every subject in the data set. It is interesting to see how subjects with higher standard deviation during epochs usually belong to the MCI group, something relatively well documented before.

Apart from a qualitative analysis, standard deviation among epochs does not give further information about subject's performance.

[13] Not always. Some subjects have only 30 or even 29 epochs.

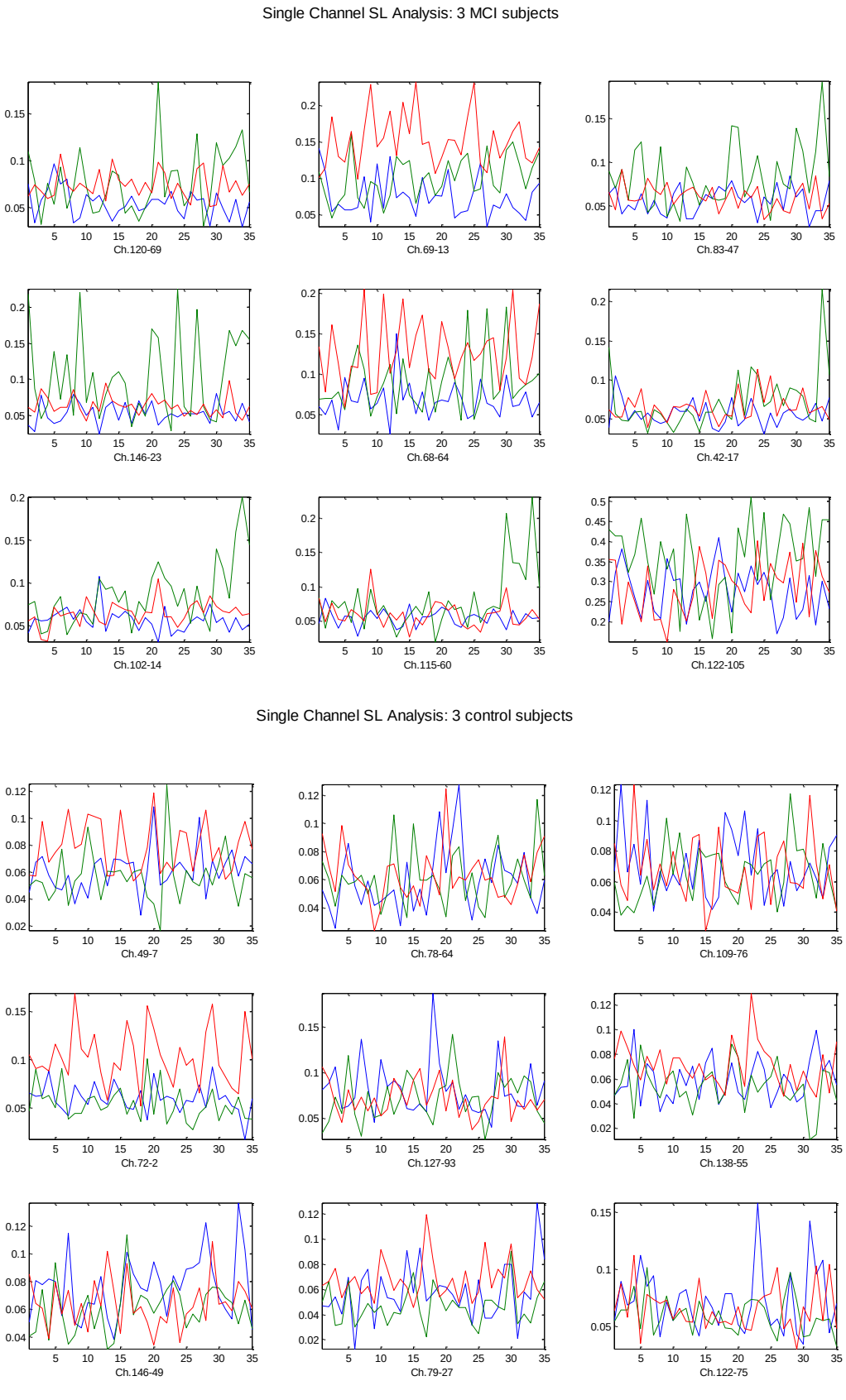


Figure 8 SL values behaviour through epochs

Dandelion graph

Another approach will be to try and map different geometrical relations between data sets in order to find out layers inside each epoch, and also among different subjects belonging to the same group.

Figure 9 show the Euclidean distance between epochs for a subject belonging to the control group.

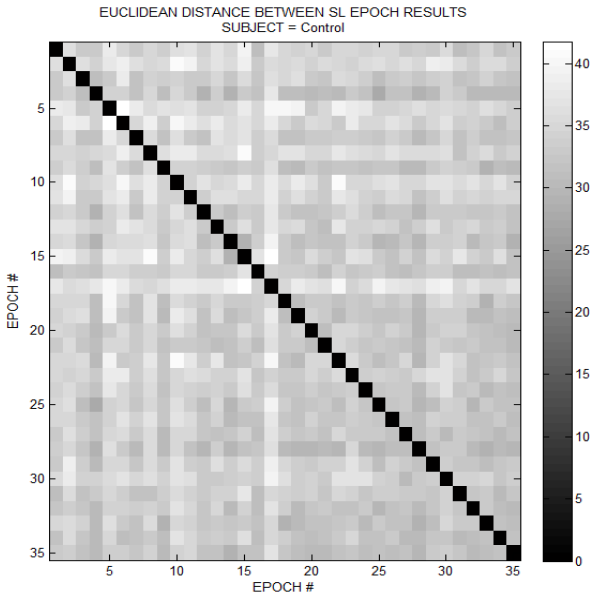


Figure 9 Euclidean distance between epochs.

However, apart from epoch #17 there is no further information which can be extracted from this representation. The following figure (totally in-house) will try to represent all the data set in a way in can be perceivable at a glance. It shows interesting to try to graph data dispersion or stability and how it changes among different subjects. This can underline possible noisy epochs, or even out layer subjects.

Euclidean distance between epochs belonging to the same subject was computed for each subject. Then a mean of all epochs will define the position of that subject, obtaining a single value for each subject. This way a measure of how dispersed epochs are among them inside each subject can be plotted as well as how dispersed are subjects between them. So for a better first sight quantitative analysis this new infographic method was developed in-house. Called Dandelion Graph because of obvious reasons, it shows all the data set at a glance.

Adopting a botanic terminology, Dandelion graph shows a first dimension of radial beaks that end up with a seed. Each seed represents a subject. The length of each beak is a way to represent several relative measures like Euclidean distance between subject means, or it could measure the same distance to the origin of the featured space.

To represent our 10878 dimensional data, the triangular inequality makes it simple impossible task. Dandelion is a proposal to try to overcome this problem.

Data patterns are not likely to behave regularly by no means. This is why different approaches, yet all adapted to the dandelion scheme, are valid. Euclidean and Chebychev distances have been tried, and linear and cosine correlation as well. Moreover, the whole work is being developed on top of basic hypothesis like if eventually data forms separable ensembles of points divided in two groups. Even though in an ungraphable 10878 dimensional ensemble space, those differentiable clouds can be sort of absorbed with a dandelion representation. The most important thing with this data representation is to try to enhance the perceptibility of each group of points.

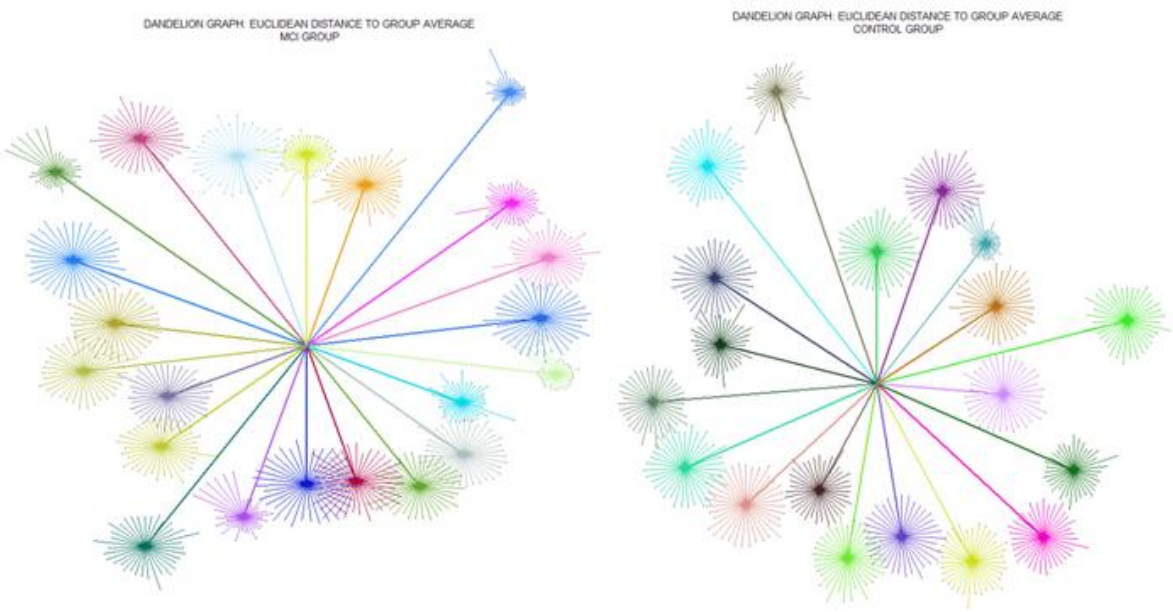


Figure 10. Dandelion graph. A condensed perspective of the data is achieved.

Figure 11 shows how when plotting the linear correlation (for normalization purposes 1-correlation is rally plotted) of every subject to the average of MCI group, members of MCI group appear to be sort of joined together. This result may be a hint for considering data set as a separable case.

A complete set of Dandelion images can be found in Appendix A.

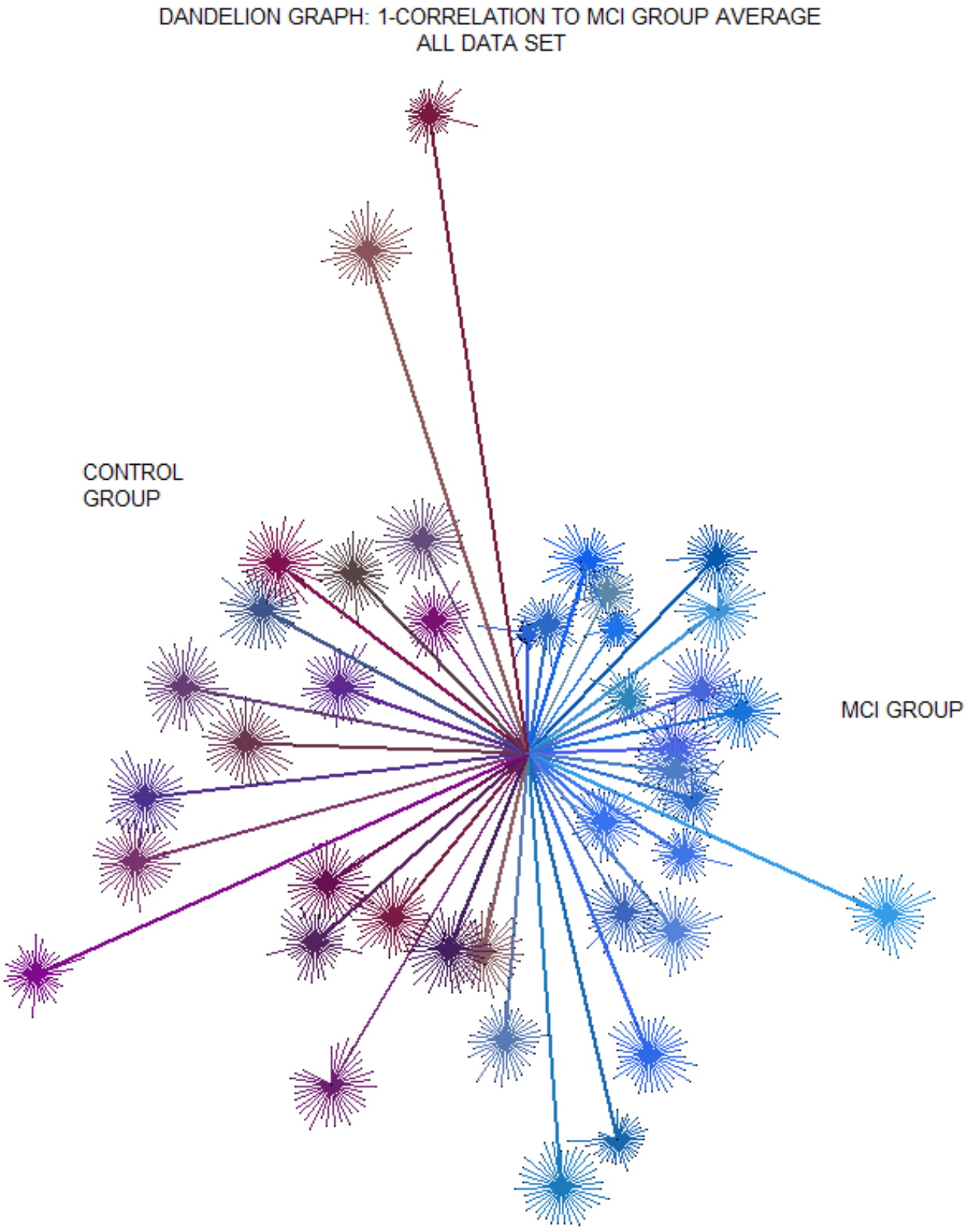


Figure 11 Dandelion Graph of complete data set.

Scaling and Equalization

It has been noted before the importance of scaling to avoid attributes in greater numerical ranges dominate those with smaller numerical ranges. By scaling usually two different data manipulations stand for scalability: (i) scale data features linearly and every feature in the same way into a range. In our case [0,1]; (ii) Scale or Equalize some features over others to let them arise as the bigger ones, as a way of underlining features eventually hidden inside raw data.

As in this case, SL values already belong to the [1,0] interval, normalization is not an issue. However, regarding the second definition, a look at the histogram of a typical SL output can be interesting.

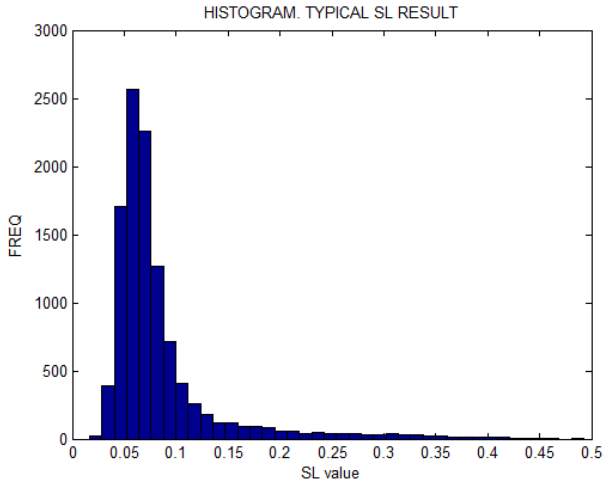


Figure 12 Typical histogram. The majority of SL values usually lay around 0,06.

Three different methods of scaling and equalization will be studied during this work, along with no scaling at all. SVM training will be exercised with the three pre-processes, in order to find the best behaviour on training results. The algorithm used for scaling were the Clahe equalization method, and

Scaling: Scaling, hear, means subtracting the mean of every one of the 10878 dimensions along all training sets. And then, adapting the histogram of the resulting vectors in order to find a flat histogram. Fig. 13 shows an example.

Equalization Method 1: This method is performed to each instance before training begins. When dealing with the issue of reducing all epochs from each subject. Clahe method is applied to each instance and then a mean average is performed. Fig 14 shows an example of its effect.

Equalization Method 2: As with previous method, it is performed to each instance before training begins. But now, first a mean average is performed to reduce dimensions from 35 epochs defining one subject, to a single one. Then, Clahe method is applied to each subject's epoch. Fig 15 shows an example of its effect.

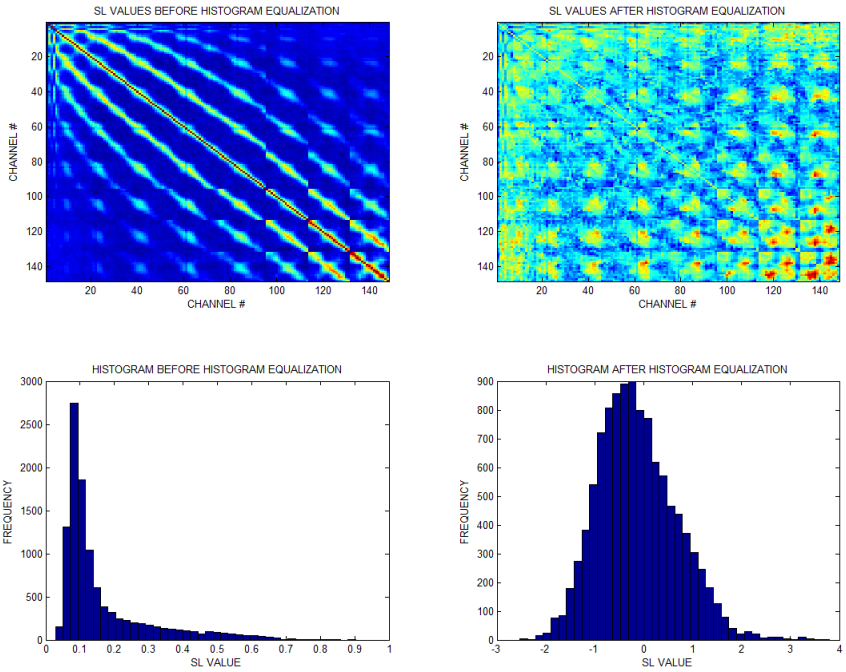


Figure 13 Effect of scaling on samples.
This is the method implemented by *svmtrain* function in Matlab.

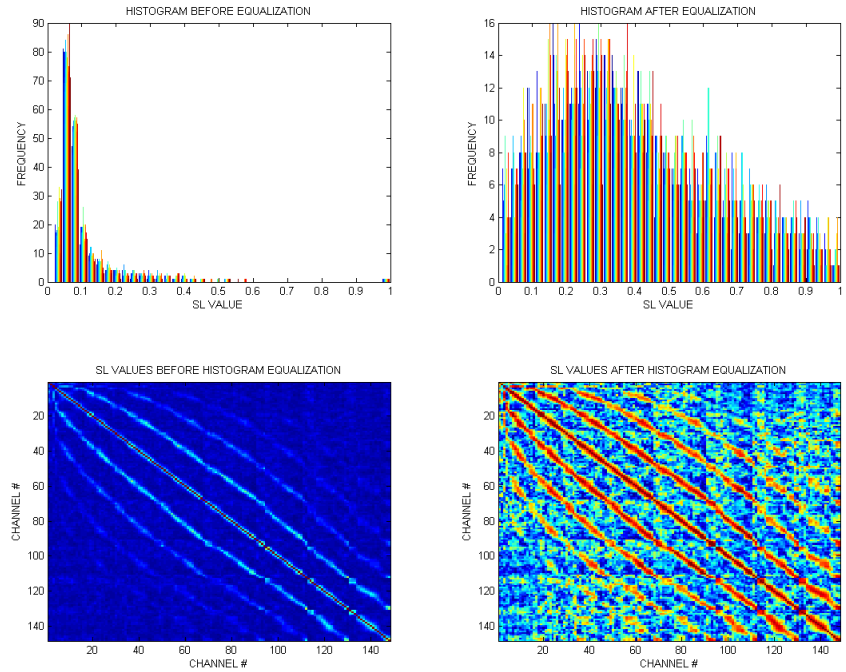


Figure 14. Equalization method#1

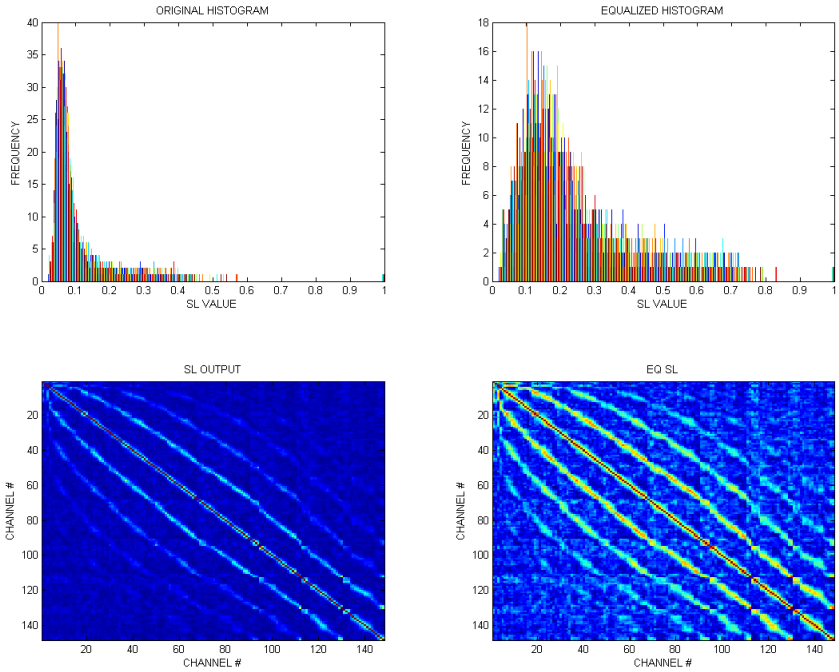


Figure 15. Equalization method #2

Clahe method operates on small regions of the SL matrix called "tiles", rather than the entire matrix. Each tile contrast is enhanced so that the histogram of the output region approximately matches a linear histogram. Neighbouring tiles are then combined using bilinear interpolation to eliminate artificial incidence boundaries. The contrast, especially in areas with homogeneous low SL levels can be hinted to avoid amplifying any noise that might be present in the image (Karel Z. 1994)

Principal Component Analysis

As standard deviation is a method for describing data spread, and mean averaging reduction might seem a little bit too drastic for reducing variables in this matter. An effort will be done in order to try and find a way of transforming data into another space where features can be more easily detected.

Principal Component Analysis (PCA) will map covariance between by pairs. A useful way to calculate de covariance between different dimensions is through a matrix: covariance matrix.

$$C^{m\times n}\left(C_{i,j}, C_{i,j} = cov(Dim\ i, Dim\ j)\right) \tag{14}$$

Every matrix represents a transformation. Within every transformation there is an initial state and a resulting state. Well, in some cases some vectors have the quality (within a defined transformation) of not being transformed at all. With the transformation described with the covariance matrix the only thing that can happen to them is to become bigger or smaller, but pointed always the same direction. These vectors are called eigenvectors. And eigenvalue is the amount by which the vector has been multiplied. Eigenvectors have the attribute of orthogonality between each other. That helps to understand under which perspective data can be analyzed and compared with the best way.

PCA is a way of identifying components in data and allows a better identification of the best way in which data can be represented in a way where differences and similarities can be raised.

Deriving the new data set will give us the original data only in terms of the new features behind the eigenvectors. This way, the coordinate's shift PCA induces over data set can make non-observable meaningful variables arise over background data.

The output of this effort can be called Eigen-Synchronization Likelihood output. It has demonstrated very useful in other machine learning applications, and here it is intended to check for better results.

Data training sets

While many papers, manuals and reference books attend the issue of pre-formatting raw data in order to maximize probability for a later satisfying result with subsequent classification and regression methods, it is not clear which technique will have best results. Therefore different approaches will be performed through this work.

Because we have two instances with only 29 epochs (in the MCI group , subject #13 and #20 and other two with 30 epochs (MCI #7 and control #... it could be reasonable to try and flatten the data and use only 29 epochs of each subject. However, in order to avoid conditioning the training phase our learning machines it can be arguable to discard 3 instances of subjects from the more populated MCI group.

Based on figure 11, MCI subjects #14, #19 and #20 will be discarded.

Mean reduction data set: This method will add significance and coherence with previous research regarding this data set. Mean averaging extracts significant features from all series of data by reducing all 35 samples of every pixel (or by meaning: every synchronization channel pair) in the sync matrix using a mean reduction.

Moreover, during memory task evaluation repetition is one of the most important factors in order to assure cognitive process under study is raised over the processes taking place in the subject's brain during the study.

Mean reduction + Equalization data set: This method will allow analyzing the sole effect of equalization.

As it has been described already, not all subjects have 35 epochs: some have 30 and the least has 29 epochs. Therefore, if an out layer procedure was to take place, according to figure 10 showing a dandelion graph, the correlations similarities the oultayer less set will be integrated by 19 MCI and 19 Control set would be integrated by subjects all with 29 epochs.

Outlayers have been considered the worst 3 epochs in each subject. This way a more homogeneous data set is obtained. It will be interesting to see the before and after of MCI and Control groups after rejecting this worst cases.

Outlayerless 29 and **Outlayerless 29 + EQ. data sets:** Both groups with 19 subjects. All subjects with 29 epochs. The majority of subjects, had their six worst epochs rejected, according to a measure of linear correlation between epochs. The linear correlation will be computed between pairs of epochs from the same subject. All measures will be averaged and ordered. Epochs with worst average will be rejected. A mean reduction is performed to the resting instances, with and without equalization being performed.

It seems reasonable to think that to reject out layers might not be a good start for a good methodology, but one important thing to note is that previously to the beginning of this work, as described in chapter 3, all subjects had 35 epochs selected from around 70 trails. Therefore, out layers have already been discarded, with no bad in very satisfactory results, up to date.

Outlayerless 19 and **Outlayerless 19 + EQ. data sets:** As with previous data set, out layers will be considered the sixteen worst epochs according to a linear correlation measure with the rest of the epochs from the same subject.

Eigen-SL and **Eigen-SL + EQ. data sets**: From a PCA analysis data is transformed and no longer is mapped in a SL-between-channels coordinate.

For ease of data use and handling each vector was wrapped into a single dimension column vector. This is, transforming a 148 by 148 symmetric matrix into a 10878 array of significant SL values, by taking into account only one half of the symmetric matrix without considering the principal diagonal. Principal diagonal carries no information at all, since SL rates at '1' the synchronizations of each channel with itself.

Training Procedure

Only 19 MCI subjects and 19 Controls subjects, sum insufficient gross data in order to perform other than cross validation methods in order to ensure the training not to fall into overfitting to the training data. Cross validation procedure can prevent the overfitting problem, and among all different possibilities, though quite expensive computationally, Leave One Out Cross Validation (LOOCV) will be chosen to perform the estimation approaches.

LOOCV Method:

i. Isolate a single sample from the rest
ii. Train with the rest of the data
iii. Quantify error with the isolated sample
iv. Repeat de previous with every single sample in the data,

The result of a training cycle will be:

Correct Ratio: Number of times the "one out" instance label prediction was correct.

Number of SV: Mean number of Support Vectors during training.

Number of Training Errors: Mean average of the number of training errors during all cycles.

Number of Iterations: Mean average of the number of iterations for the dual optimization problem, given by expression (6), to find a solution.

Five main kernel functions will be tested during training: Linear, Quadratic, Polynomial, Radial Basis Function and Multilayer Perceptron.

Enhanced Recursive Feature Extraction

In order to find out which of the 10878 dimensions tend to drive more the label decision in one way or another. And by other means, try to underline which dimensions are meaningless to the process of labelling a subject as MCI or HE, in terms of SL connectivity.

In this work, data could be considered high-dimensional low-sample size data (HDLS). And for this type of data, it is particularly important to reduce dimensionality due to the difficultness of obtaining new samples and the high dimensionality of each one.

Recursive Feature Elimination (RFE) is a method for selection of relevant features (dimensions) for a defined embedded method. In our case it is particularly suitable for applying over SVM optimization results, due to the way in which decision function is build. It lays in SVM condition described in equation (7) As well as in the good generalization performance inherent to the SVM.

As observed with PCA analysis, one big problem with weak components elimination is that relevant information usually relies on nuances that show off as low qualified relations by means of pondered relations. This is why some times subtracting low rated statistical differences between datasets can harm classification efficiency and generalization "capacities"[14].

RFE tries to improve generalization performance by removing the least important features, but it may happen that low rated features tend to have crucial role in generalization performance. Thus, removing weak features (sometimes redundant) can degrade classification result.

Enhanced Recursive Features Elimination (EnRFE) will improve RFE performance by rating weak features when they demonstrate useful for classification while combined with other features. EnRFE recursively removes features at each step and re-ranks remaining features by re-training SVM based on the remaining features.

[14] The meaning of *capacity* refers to the idea of a "capacity to learn" as expressed by Vapnik (Vapnik 1998).

RESULTS

Equalization method analysis

As SL values matrixes are a 148 by 148 pixels image, this work can benefit from image processing techniques. If the histogram of the image is analysed the first thing that arises is that there is a much higher frequency of low SL values. This is a typical case where a nonlinear enhancement of the "image" can be done. But until now, the doubt might rise whether a non linear transformation might be altering and deteriorating the capacity of our classification.

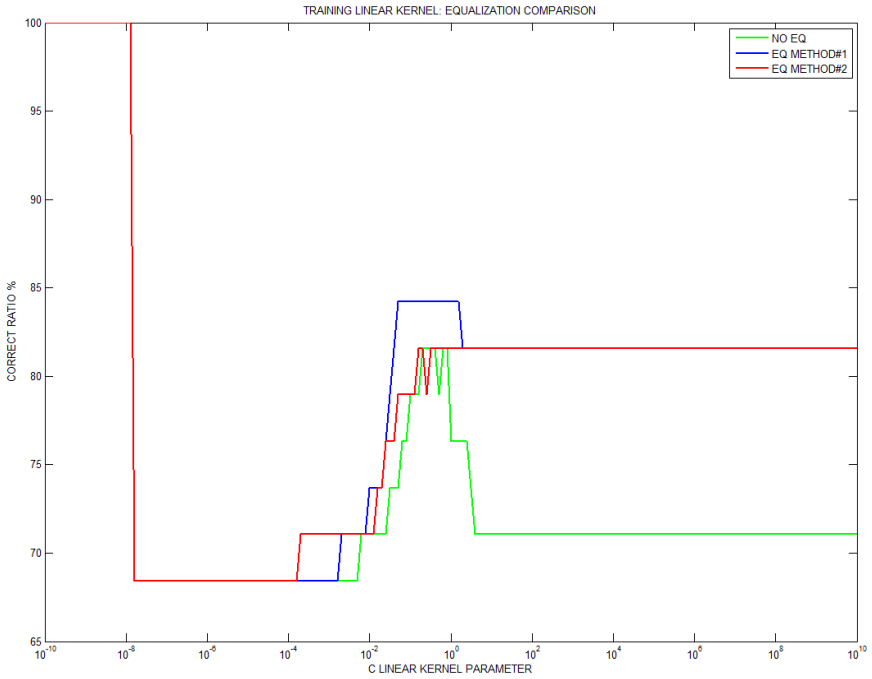


Figure 16. Linear Kernel training. Equalization comparison

Figure 16 shows different results for a LOOCV training of a linear kernel function. Linear Kernel can only be trained with the C SVM parameter which defines penalization for training errors. This is, when C is low, (always $C \geq 0$) training errors suffer less penalization, while solving the dual optimization problem. The training set used was the mean reduction data set.

In this case it was proved how equalization method #1 improved training results, while equalization method #2 had no improvement in training results. As results show, scaling by method 2 will deteriorate results by approximately a 5-10%. And as one of the main objectives in this work is to develop an efficient methodology for analysing SL analysis with SVM, a method with computational cost and no benefit in training output, will be discarded. The maximum training Correct Ratio for method #1 is 84.21% but this is a matter of discussion later.

It must be pointed out over fitting results for Correct Ratio, with C bellow 10-4. This results typically appears when training optimization algorithm allows too many training errors. This way, a good way to check for reliability in the results is to plot Correct Ratio results along with the number of Training Errors during training.

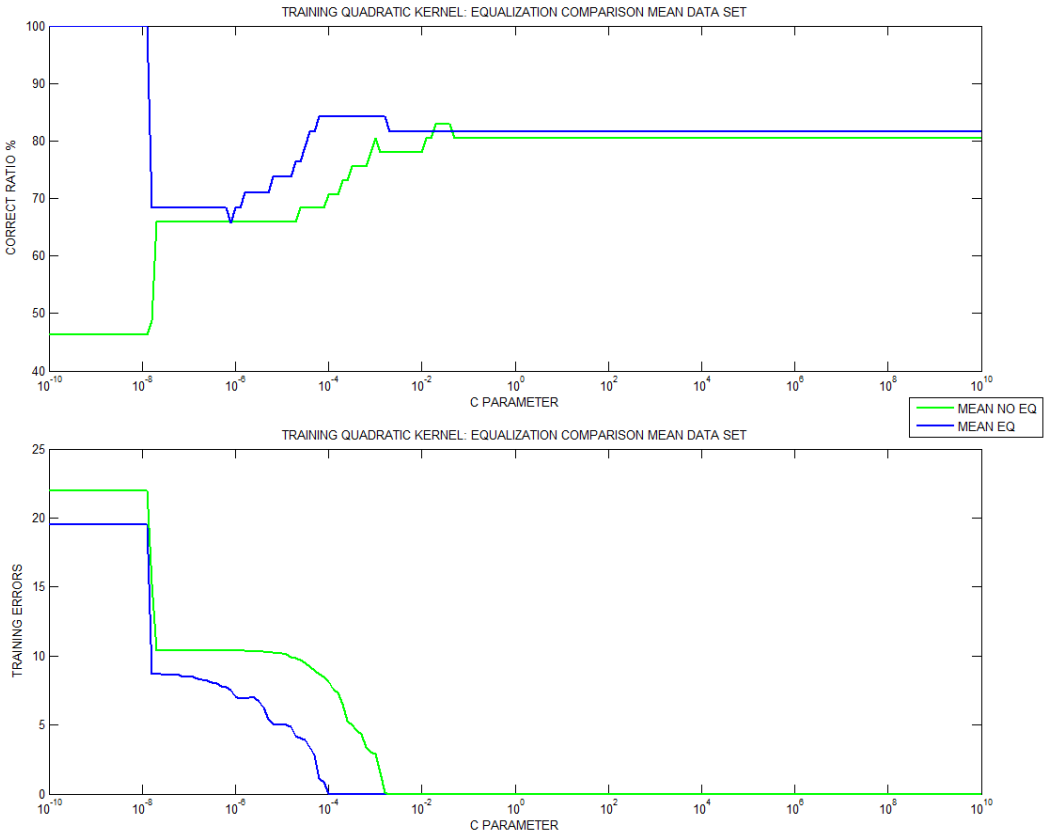


Figure 17. Quadratic Kernel training analysis.

As shown in previous figure, up to 22 from a training set of 37 (training errors is given as a mean average along a complete LOOCV cycle), were badly classified by the trained SVM.

For now on, method #2 will no longer be used.

Scaling analysis

The next two figures show different results for training on different data sets while comparing different methods of scaling. This result has been continuously repeated along all data sets and along all different kernels, therefore by this result, it is considered proved to produce worse results whenever data is already enclosed between 0 and 1 as in, our case.

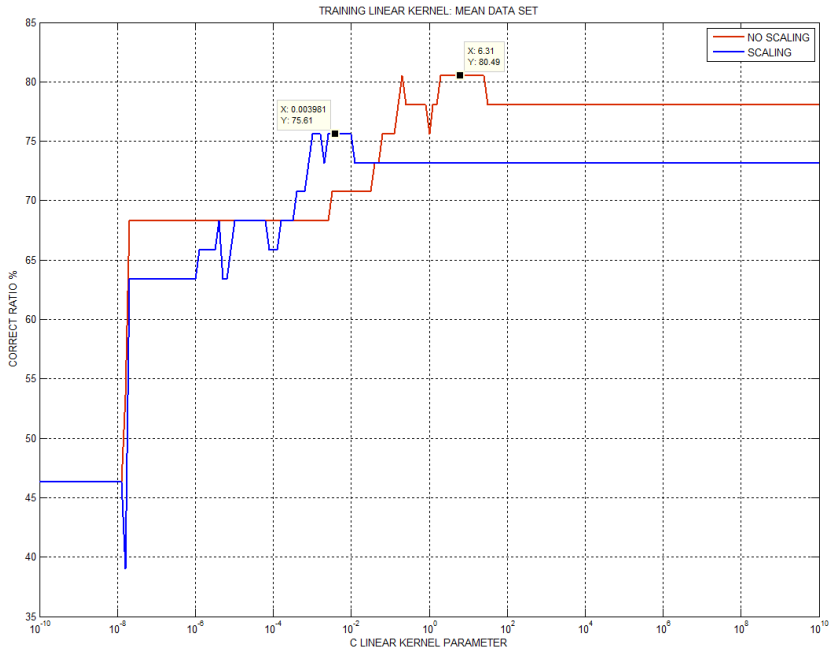


Figure 18. Linear Kernel training: Scaling comparison

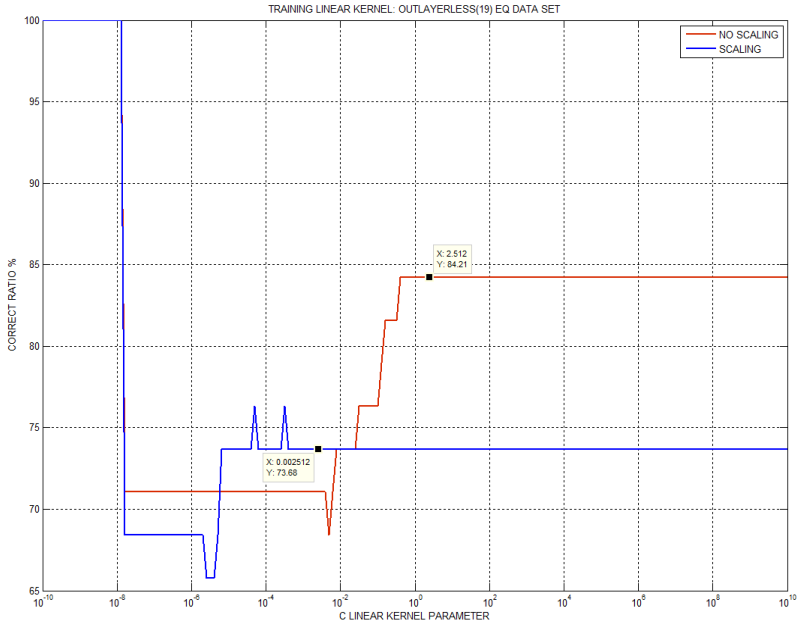


Figure 19. Linear Kernel training: Scaling comparison

Later, maximum Correct Ratio results will be collected in table 3, but before attending these results, it is interesting to attend the overlearning results presented in the second and third plots. As it can be clearly seen, a 100% classification is obtained for C parameters bellow 10^{-4}. The next figure will show what is happening during the training.

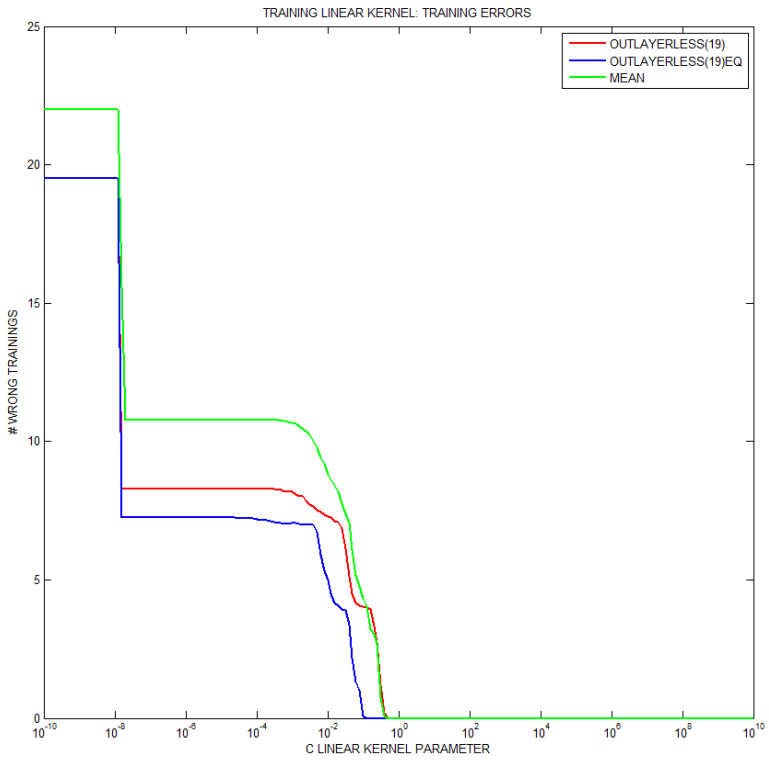


Figure 20. Linear Kernel training comparison among different training sets.

As shown by previous figure, training is done with 37 samples each LOOCV turn. Therefore to have a high number of training errors, invalidates the 100% Correct Ration obtained. But to have training errors does not necessarily mean a bad training has occurred. For instance, a non separable case will not be able to converge into a maximum margin hyperplane as solution of the optimization problem, if no training errors are allowed.

For assuring as much as possible a good generalization performance of the SVM, from now on we shall consider always low training errors rates.

Returning to the scaling problem, we shall not consider training correct ratios bellow C=1, for this particular case, and therefore scaling is confirmed to deteriorate correct ratio results by approximately 10%. This behaviour was confirmed along all training exercises developed while doing this work.

PCA results analysis

This result has been widely used with other machine learning methods. However, it is the *eigen Likelihood Synchronization matrix* does not seem to rise or discover hidden features in data that might make SVM training output improved correct ratio results.

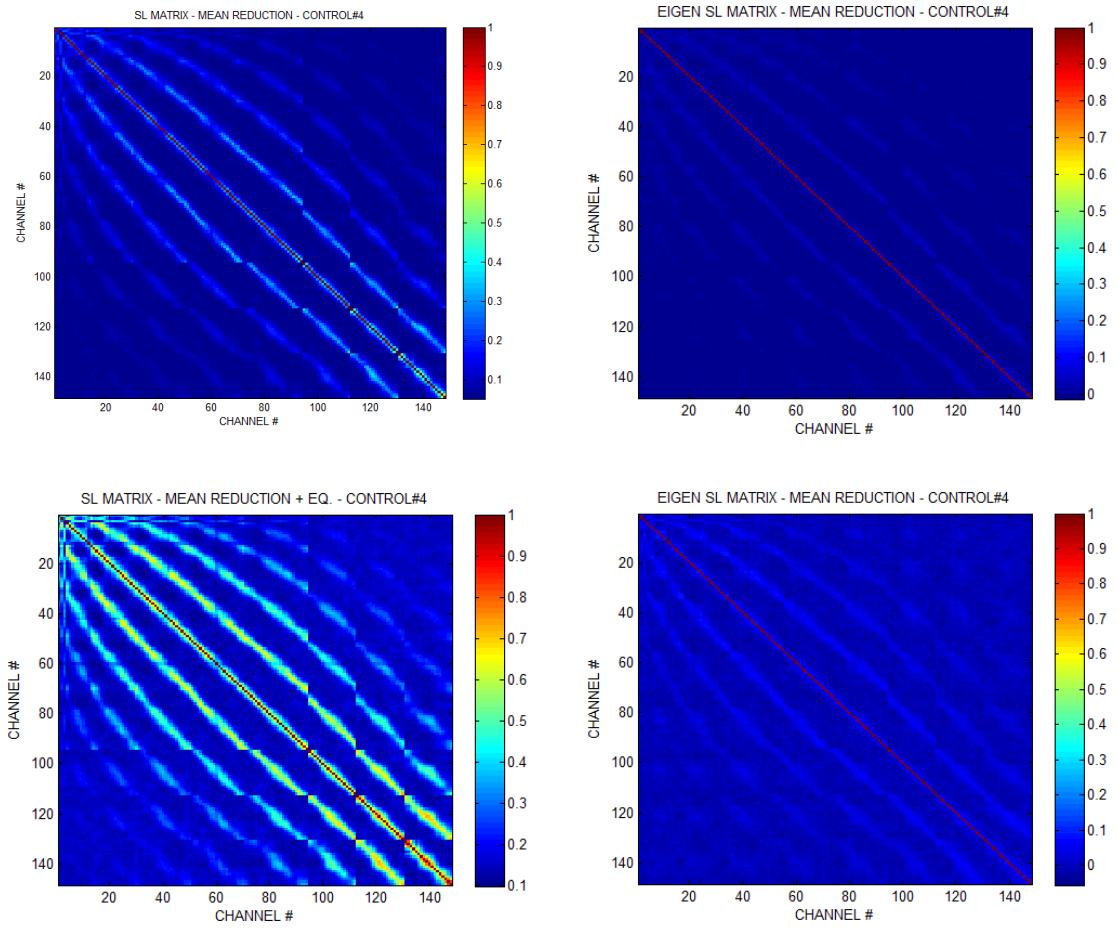


Figure 21. PCA effect on SL matrixes.

As it will we clear during the next pages, Radial Basis Function kernel produces stable results. After evaluating where RBF kernel parameters (sigma and C) produced better results, the eigen-SL data set was tested obtaining a 76.32 % Correct Ratio, at maximum. Next figure shows different Correct Ratio values obtained during grid training around this localized area.

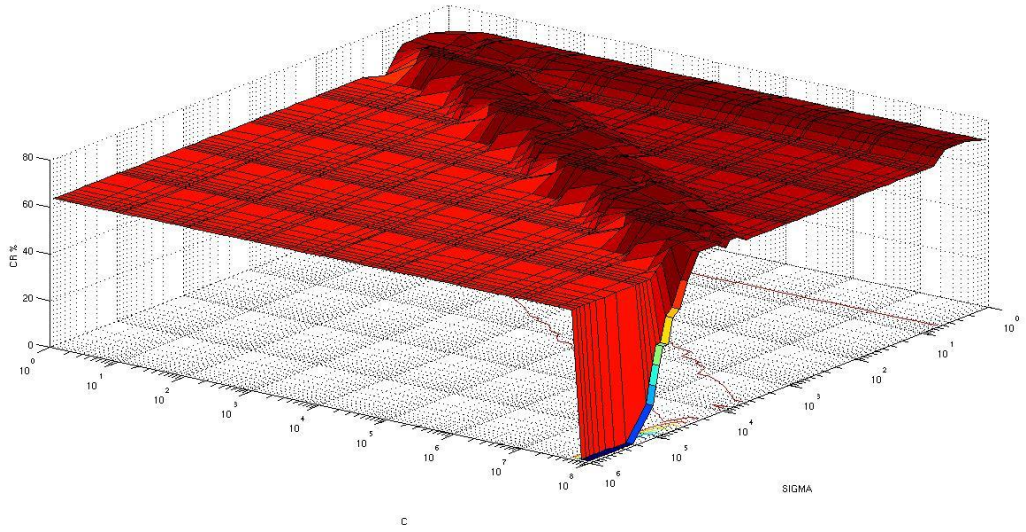


Figure 22. RBF Kernel on PCA's EigenSL matrix data set.

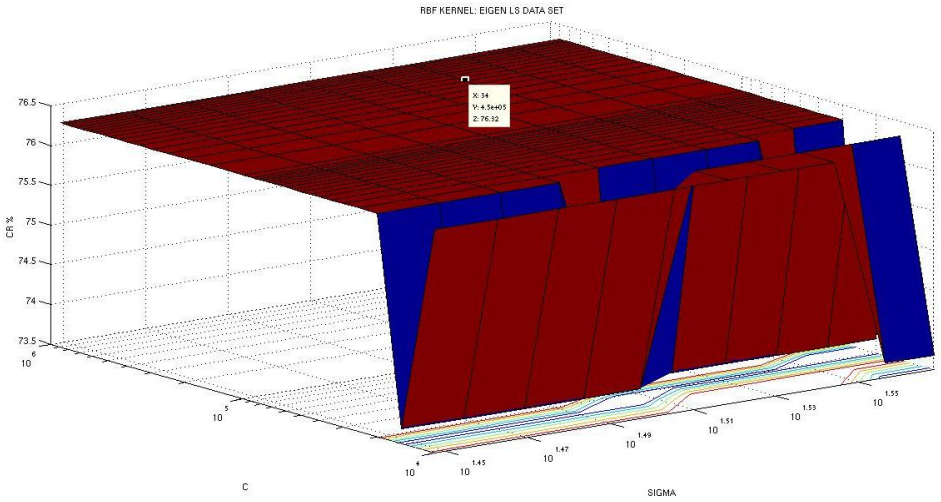


Figure 23. RBF Kernel on PCA's EigenSL matrix data set: closer look up.

Linear Kernel

A grid search for optimum kernel parameters was performed for all six remaining data sets. Next figures plot results.

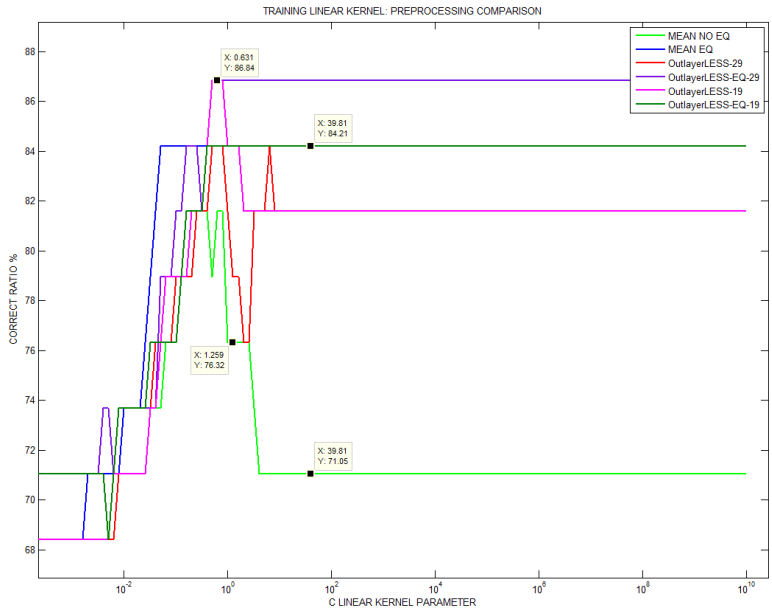


Figure 24. Linear Kernel training for all data sets.

Surprisingly Outlayerless 29 +EQ data set, produces better results than other data sets with more outlayers rejected. As with previous results shown in figure 24. Maximum Correct Ratio values are considered stable on minimum Training Errors states.

Quadratic kernel

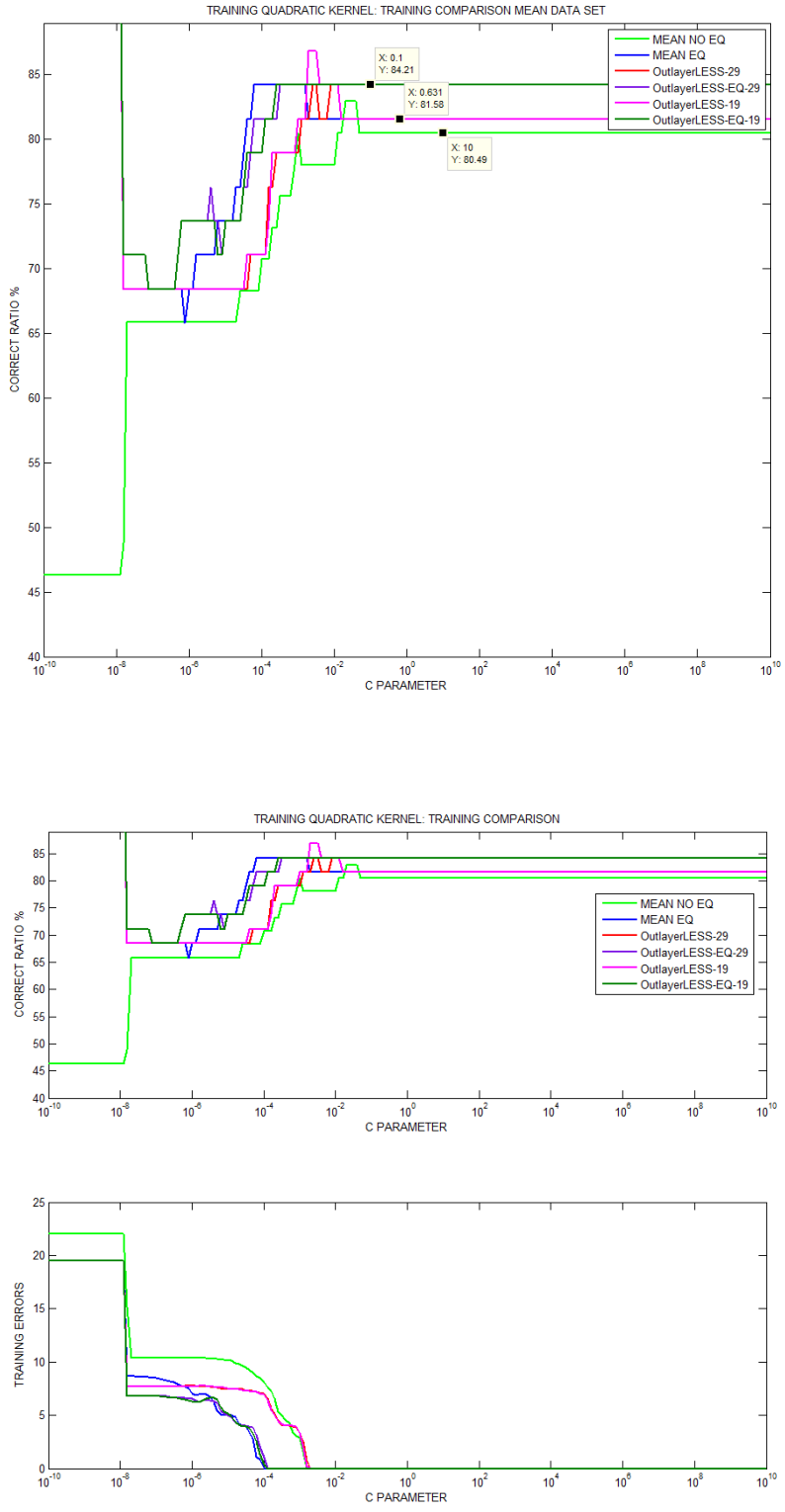


Figure 25. Quadratic Kernel training results

Radial Basis Function Kernel

A grid extensive parameter search was performed in order to find the best performance of the SVM training.

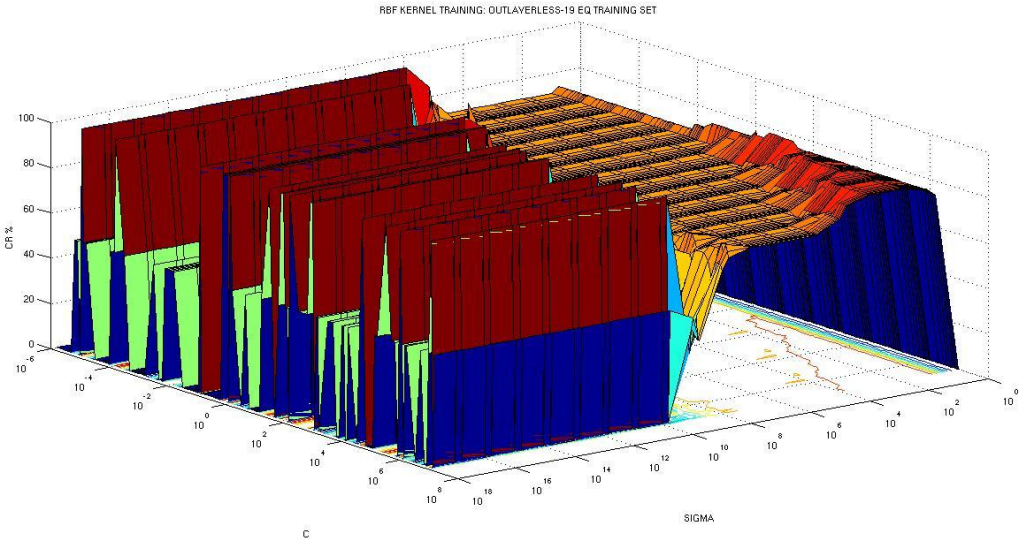


Figure 26. RBF Kernel training. Wide grid training.

Last figure shows how wide areas of the result belong to overfitting regions with 100 % Correct Rates yet with very high Training Errors. While not plotted, for negative sigmas, no more than 70% correct ratios was obtained with this kernel. Next figure shows another perspective of the resulting CR.

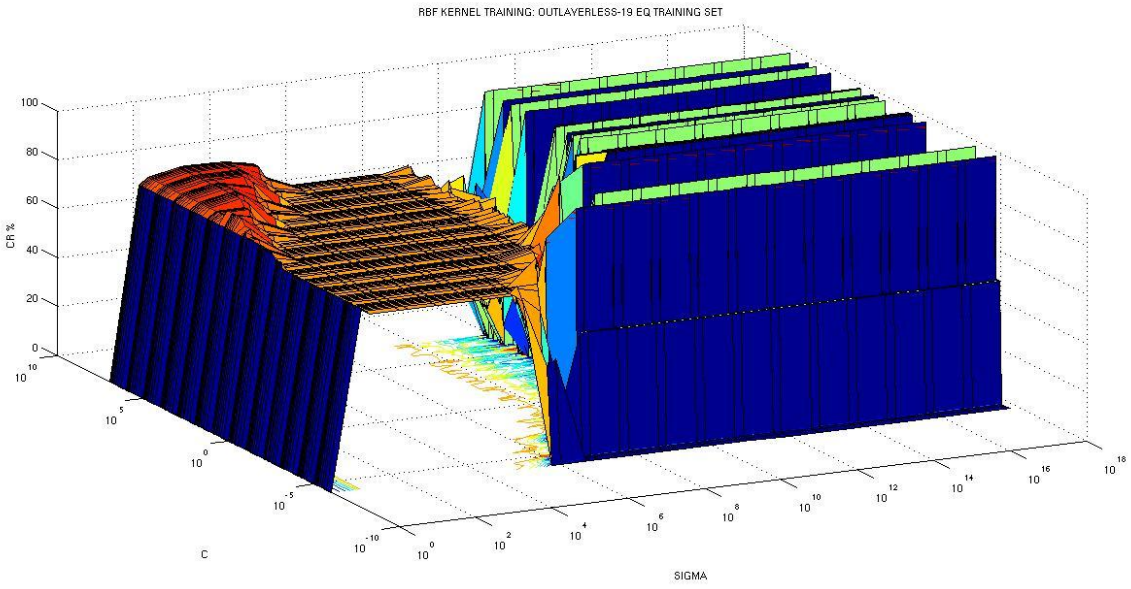


Figure 27. RBF Kernel training. Wide grid training. Another perspective.

To ensure the top left distinction is a good result area, we can plot Training errors, from a similar perspective. Obtaining no training errors for that zone.

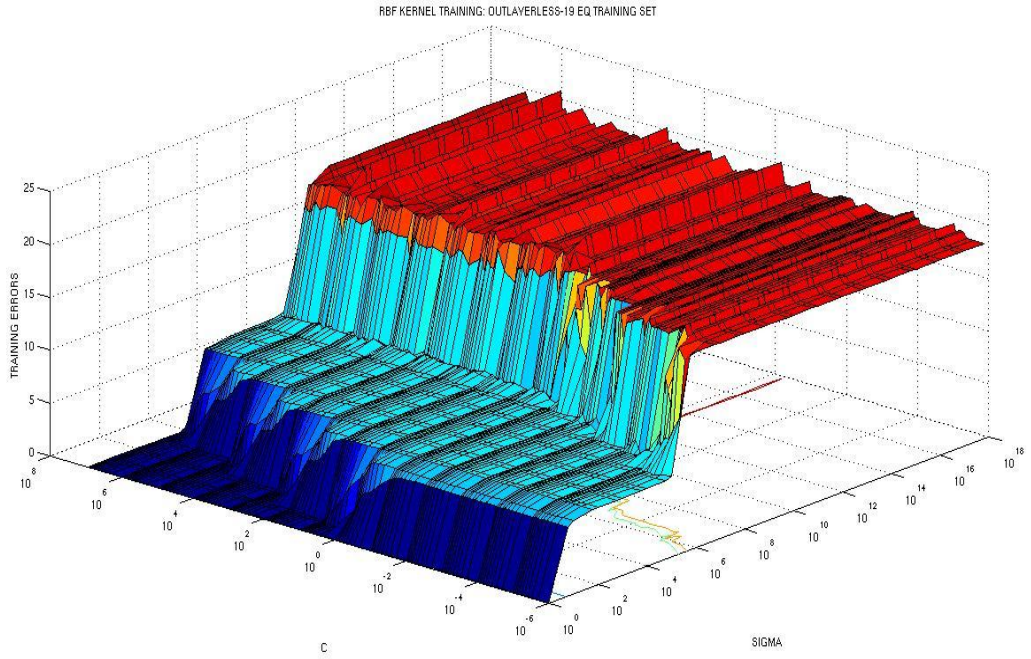


Figure 28. RBF kernel training errors.

And here a closer look up shows a maximum CR of 84,21 %. But to be sure, it is necessary to narrow the search grid around the maximum area.

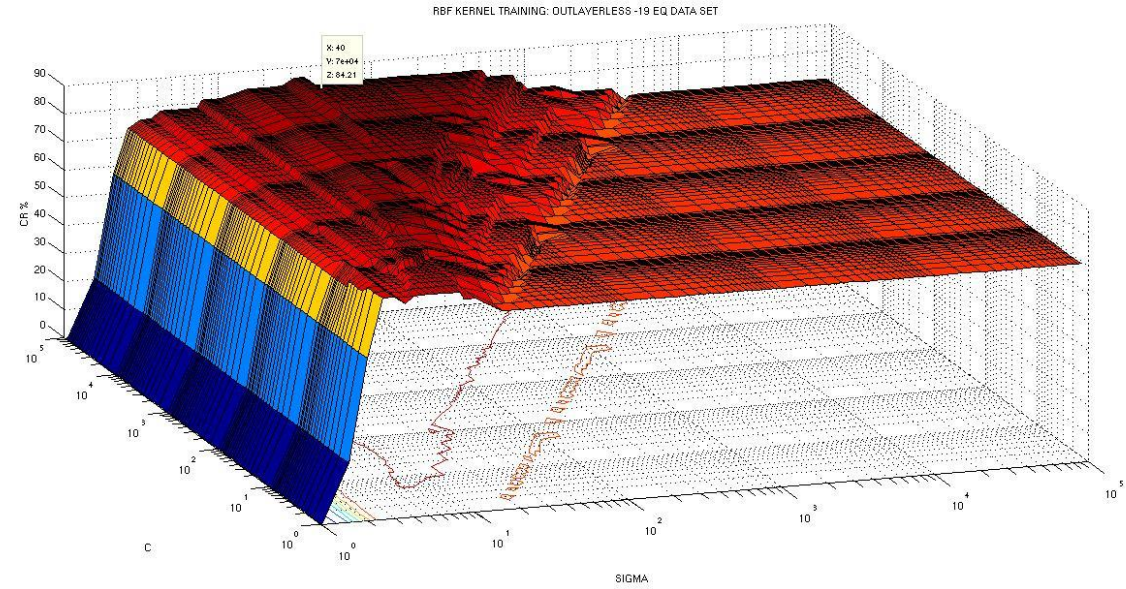


Figure 29. RBF kernel training. Maximum Correct Rate values.

A possible way to obtain an idea of how generalization performance is going to perform, when new data is presented to de SVM, is to plot the number of Support Vectors. In this case, the maximal area happens to coincide with a minimum of SV number, area.

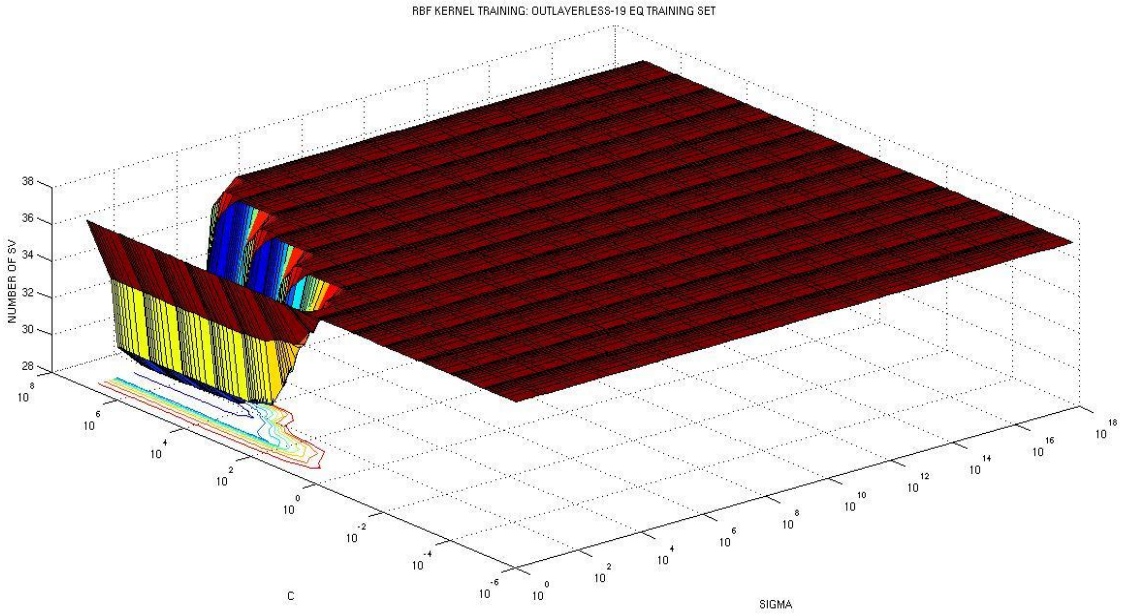


Figure 30. RBF Kernel training. Number of SV.

When kernel training is narrowed, the maximum appears quite stable. A 86.84 % Correct Ratio.

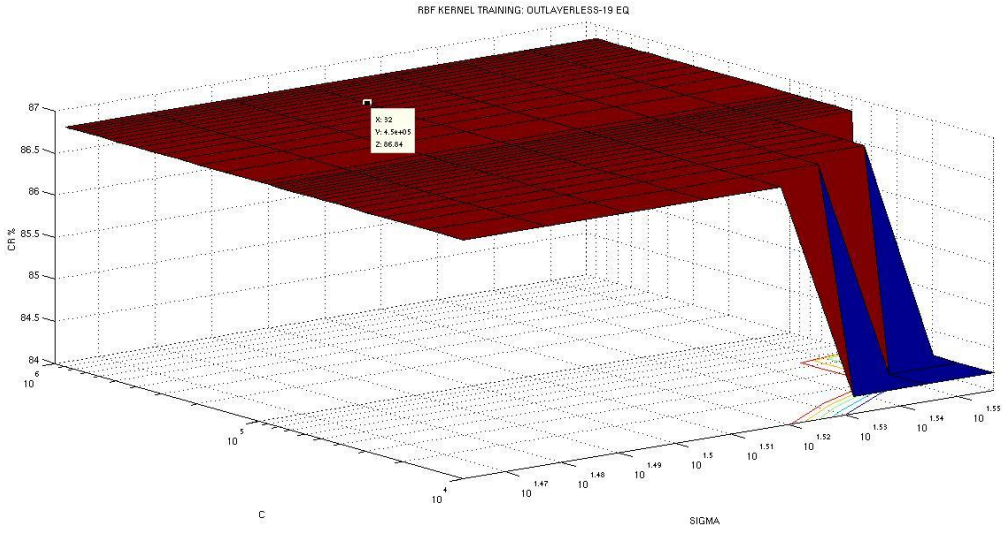


Figure 31. RBF Kernel training. Close look at maximum.

Polynomial Kernel

As like the RBF Kernel function Polynomial Kernel training was performed by doing a grid search as wide as possible. Obtaining a 86.84% of Correct Ratio.

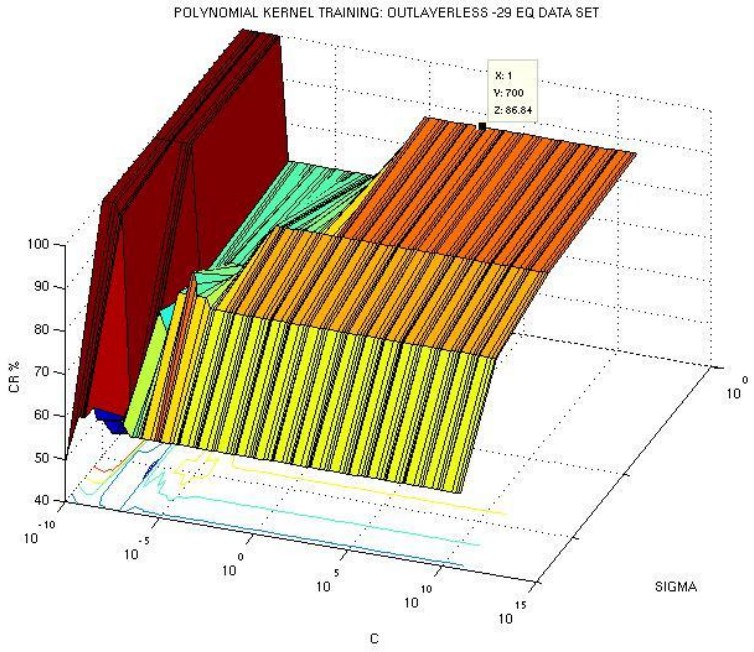


Figure 32. Polynomial Kernel training.

With a stable number of support vectors.

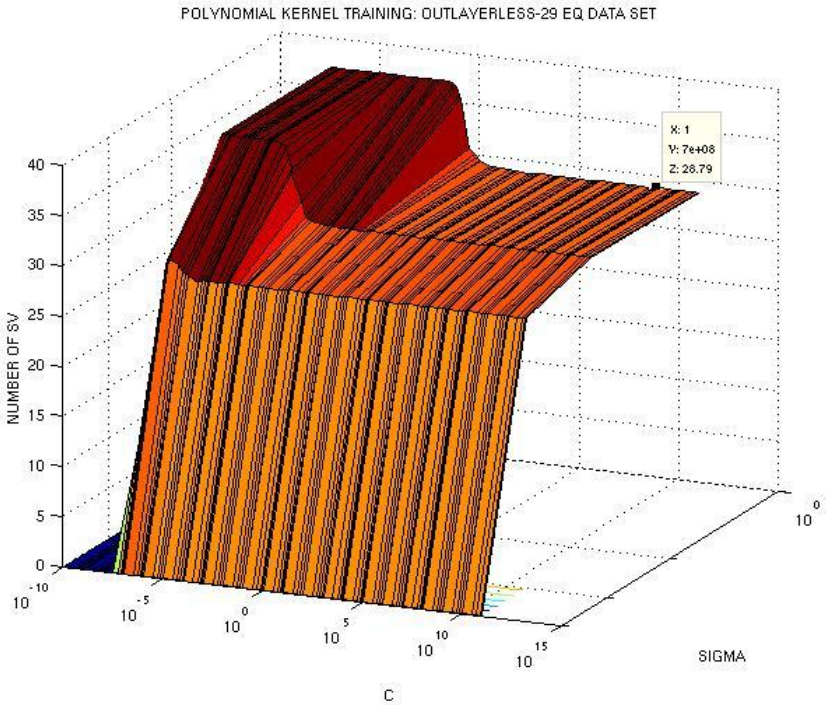


Figure 33. Polynomial kernel training

Multilayer Perceptron Kernel

For this kernel, results did not went further than 50 % Correct Ratio, and Training Errors did not went bellow 15. Therefore, MLP kernel was not considered for any further analysis.

Comparative Analysis

Next table resumes all training performances, for all considered data sets. All results where verified to occur while having no training errors during the training process. Which means, not a single training error, during all 38 LOOCV cycles.

	KERNEL TRAINING: CORRECT RATIO [%]			
Type of Data Reduction	***LINEAR***	***POLYNOMIAL***	***QUADRATIC***	***RBF***
Mean	**76.32**	**81.58**	**80.49**	**80.49**
Mean + EQ.	**84.21**	**81.58**	**81.58**	**81.58**
Outlayerless 29	**81.58**	**81.58**	**84.21**	**81.58**
Outlayerless 29 + EQ.	**86.84**	**86.84**	**84.21**	**84.21**
Outlayerless 19	**81.58**	**81.58**	**81.58**	**81.58**
Outlayerless 19 + EQ.	**84.21**	**84.21**	**84.21**	**86.84**

Table 3. Comparative Analysis of different kernel trainings

As mentioned earlier, training with Multilayer Perceptron (MLP) Kernel was also checked, with no stable result for any value during the training. The normal value for CR during training was 50% with 15 training errors at best.

EnRFE

Enhanced Recursive Feature Elimination was applied to the Outlayerless19 + EQ data set and RBF kernel function tuned with the kernel arguments that produced the maximal 86.84% result. Following the algorithm procedure, all 10878 dimensions were sorted in order of importance and the least significant feature was removed each iteration.

Each iteration the SVM was re-trained producing a Correct Ratio Output which can be seen in next figure.

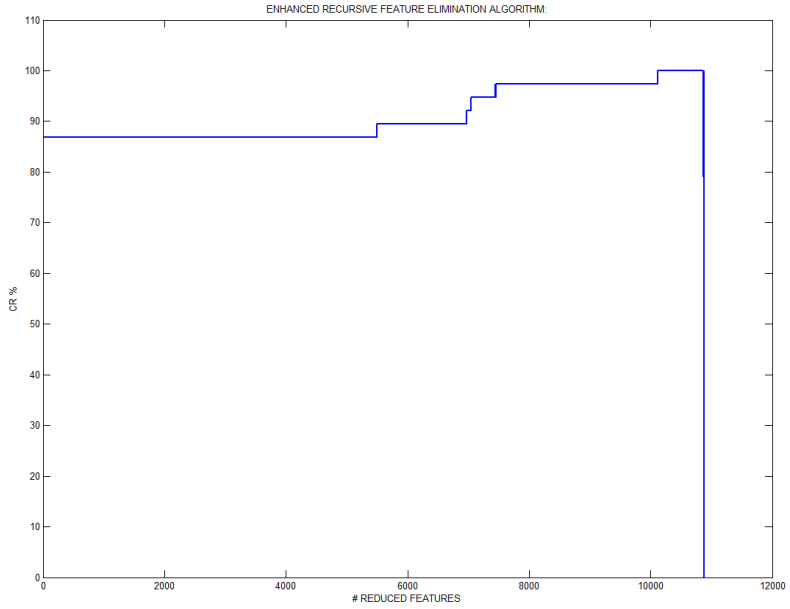


Figure 34. EnRFE wide view

With a closer look up plotted in the next figure.

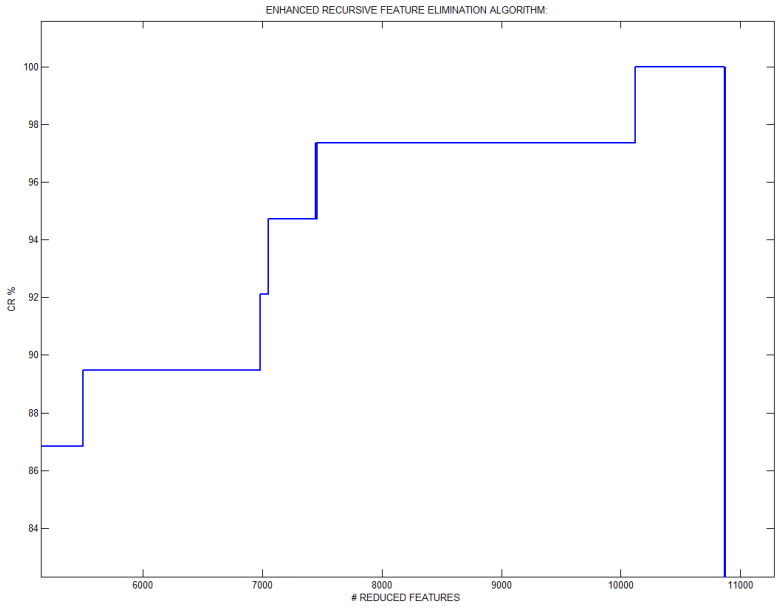


Figure 35. EnRFE closer look up.

Though, as it can be seen, a 100 % correct ratio is obtained, the information which can be obtained here relay more in what SL values produce higher weights in the decision, rather than the decision by its own.

Next figure plots the 3% more important SL values for the trained SVM in order to make a prediction. Consider the nose of the patient to be right beneath the image title.

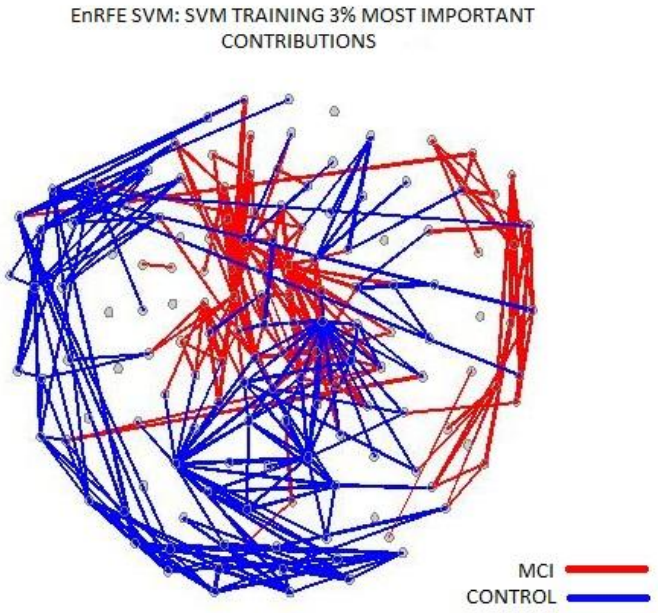


Figure 36. EnRFE 3% most important components

Conformal Prediction

Previous results show how three different kernels obtained a maximum output of 86.84% of correct classifications with a LOOCV procedure. Next table shows raw result for a Conformal Prediction execution, for these three mentioned results.

1. LINEAR: SVM with Linear Kernel for Outlayerless29 + EQ. Data set.
2. POLYNOMIAL: SVM Polynomial Kernel for Outlayerless29 + EQ. Data set.
3. RBF: SVM Radial Basis Function Kernel for Outlayerless19 + EQ. Data set.

If a confidence level of 95% is to be decided, all p-values bellow or equal to ε=0,05 can be equalled to zero.

	N# PREDICTION ERRORS	
	CONTROLS	***MCI***
LINEAR	3	4
POLYNOMIAL	3	4
RBF	3	4

Table 4. #n Prediction Errors for three best results

Both Polynomial and Linear kernels fail to predict correctly the labels of the same controls and MCI. And they all end up with a Correct Prediction Ratio of 81.57 %.

			CONFORMAL PREDICTION							
			LINEAR KERNEL			**POLYNOMIAL KERNEL**			**RBF KERNEL**	
			CREDIBILITY	CONFIDENCE		CREDIBILITY			CREDIBILITY	CONFIDENCE
CONTROL GROUP	1	C	24%	100%	C	24%	100%	C	18%	100%
	2	M	42%	95%	M	42%	95%	M	37%	95%
	3	C	68%	84%	C	68%	84%	C	66%	82%
	4	M	11%	89%	M	11%	92%	C	11%	92%
	5	M	42%	61%	C	39%	61%	M	39%	63%
	6	C	13%	100%	C	13%	100%	C	16%	97%
	7	C	76%	87%	C	76%	84%	C	97%	84%
	8	C	47%	68%	C	45%	63%	C	45%	63%
	9	C	42%	71%	C	42%	74%	C	47%	68%
	10	C	37%	92%	C	37%	97%	C	26%	92%
	11	C	34%	89%	C	34%	89%	C	39%	89%
	12	C	29%	100%	C	26%	100%	C	29%	100%
	13	C	97%	95%	C	97%	95%	C	97%	95%
	14	C	63%	89%	C	63%	89%	C	55%	87%
	15	C	50%	53%	C	53%	50%	M	53%	47%
	16	C	21%	82%	C	21%	82%	C	24%	82%
	17	C	26%	87%	C	29%	89%	C	34%	87%
	18	C	97%	58%	C	97%	58%	C	97%	55%
	19	C	97%	79%	C	97%	79%	C	97%	68%
MCI GROUP	1	M	97%	97%	M	97%	97%	M	97%	97%
	2	M	45%	92%	M	47%	92%	M	42%	82%
	3	M	55%	61%	M	58%	61%	M	61%	47%
	4	M	61%	87%	M	55%	87%	M	58%	82%
	5	M	97%	92%	M	97%	92%	M	97%	92%
	6	C	45%	97%	C	42%	97%	C	53%	97%
	7	C	18%	100%	C	18%	100%	C	32%	100%
	8	M	74%	97%	M	71%	97%	M	74%	97%
	9	M	53%	89%	M	50%	89%	M	63%	92%
	10	M	66%	87%	M	66%	87%	M	71%	89%
	11	M	71%	95%	M	74%	97%	M	68%	95%
	12	M	97%	97%	M	97%	97%	M	97%	97%
	13	M	58%	92%	M	61%	92%	M	50%	92%
	14	M	16%	92%	M	16%	92%	M	21%	89%
	15	M	97%	97%	M	97%	97%	M	97%	97%
	16	M	32%	89%	M	32%	89%	M	32%	89%
	17	M	97%	92%	M	97%	92%	M	97%	92%
	18	C	71%	92%	C	71%	89%	C	71%	92%
	19	C	37%	82%	C	37%	82%	C	61%	87%

Table 5 Conformal Prediction Credibility and Confidence values.

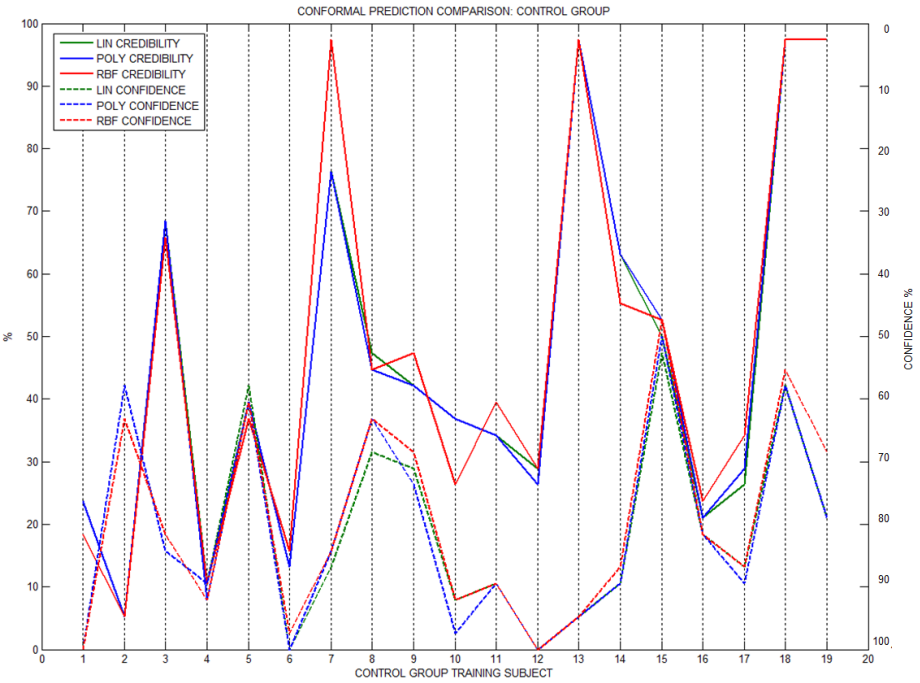


Figure 37. CP for Control Group

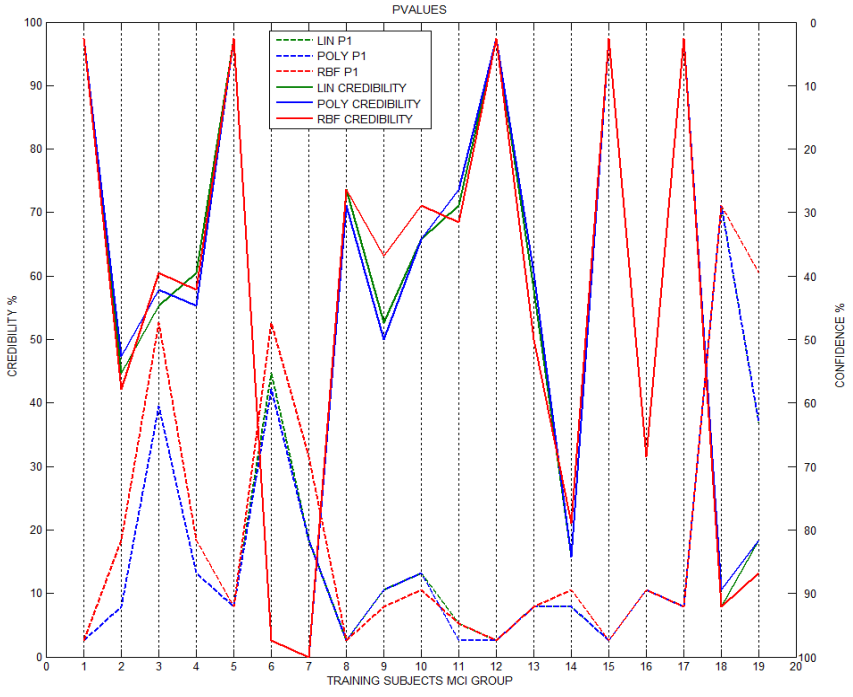


Figure 38 CP for MCI Group

Fig 37 and 38 show confidence and credibility measures for all training subjects belonging to the control group and MCI group. It seems particularly interesting how all three methods consider subjects #6 and #7 from MCI group, badly initially classified, by means of differences and similarities with other training subjects.

CONCLUSIONS

While training, it is the number of hyperparameters which influences the complexity of the model selection. Usually polynomial kernels have non-linear hyperparameters, while it is not the case for the RBF kernel (Hsu et. al, 2010).

It has been reasoned why mean averaging between different epochs must ensure a correct moderated measure of a neuropsychological process, rejecting this way transitory noises not present in a significant amount of epochs. This is, the mean can also be considered a feature extractor (Bajo et al, 2010). In this work different out layer rejecting procedures were performed and compared. Surprisingly, rejecting 16 out layer epochs, out of a total of 35 performed worse than rejecting 6, with two of the three best results. This result might rely on the solidness of LOOCV procedure.

Regarding SVM, one of the main issues has been the choice of a kernel function. As this work pretends to help developing a standardized technique for MCI patients classification with reliability, perhaps a deeper analysis in terms of scoping other kernels could possible improve not only classification accuracy, but also reliability of the classification.

As shown in previous results, three different kernel LOOCV trainings have found the same 86.84 % correct classification ratio. While with CP the correct ratio has fallen to 81.57 % however giving confidence measures for each single prediction.

The choice of the kernel parameters has also been the mayor cause of computational cost along this work. The majority of procedures and functions were executed almost straight forward with a reasonable regular computer, however the need for a grid search of kernel parameters usually turned training into a several hours (even days with the Radial Basis Function kernel – mainly because it depends in 2 parameters). Moreover, linear kernel LOOCV training is one order of magnitude faster than the other two kernels: polynomial and radial basis function. As noted, this is mainly due to the fact that linear kernel only depends on the penalty parameter C. Still, if a closer look at each training cycle is taken, quadratic optimization procedures for linear kernel approximately doubles the number of iterations towards a satisfactory solution for polynomial and radial basis function kernels.

Generally RBF kernels seem a reasonable first choice. This kernel usually non-linearly maps samples into a higher dimensional space, while for instance, linear kernel can handle cases where the relation between class labels and attributes is linear.[15]

[15] The fact that linear kernel has been able to find a solution might not necessarily mean the problem is linearly separable. The fact that quadratic optimization for linear kernel takes approximately twice the number o iterations it takes for the other kernels may give a hint over the possibility of linear kernel mapping data in a higher dimensional feature space, with linear solution. And more importantly: one of SVM most important advantages is that while working with kernels, there is no need to know how the feature space is, at all. SVM rely on Mercer's conditions applied to the selected kernel, and can perform the quadratic optimization nicely with the dot product implemented as a kernel function: which is known as the dual problem.

Attending a qualitative measure of confidence in the prediction. However it has been proved (Burges, 1998) that not always the number of support vectors of a trained SVM is directly related to the generalization risk, it might be interesting to point show this number for comparison with CP measures. Table 6 shows how polynomial kernel training throws slightly smaller number of support vectors. Apart from the fact that all three numbers are quite high, related to the number of training samples (38-1 in LOOCV).

	AVERAGE OF #SV
LINEAR	28.79
POLYNOMIAL	26.89
RBF	28.31

Table 6 Comparative analysis of the number of support vectors at the maximum correct ratio, training points, averaged through a complete LOOCV cycle.

Concerning Conformal Prediction results, it is first noticed how the same subjects are both classified with maximum confidence with either method, or how others are even (apparently) badly initially classified. RBF kernel shows slightly improved results in classifying MCI group members while obtaining higher credibility in correct predictions. However it also obtains lower confidence in the prediction. Next table shows mean average of credibility and confidence in each prediction for three proposed kernel methods.

	Conformal Prediction	
	Credibility	***Confidence***
LINEAR	55.40 %	87.05 %
POLYNOMIAL	55.26 %	87.12 %
RBF	57.13 %	85.66 %

Table 7 CP differences for three best trainings for three different kernels.

Though mean average might not show real performance due to high fluctuation in credibility values, as seen in figures 37 and 38, RBF kernel submits the better performance.

Apart from Likelihood Synchronization, a group of different algorithms take part in studies regarding MEG, EEG or even fMRI functional connectivity studies. With not to much effort, a wider study could take advantage of the same methods developed here, and apply them to different synchronizations data output. Cross correlation, Coherence and Partial Coherence, Granger Causality, Mutual Information, Phase Synchronization or Generalized Synchronization are a group wide enough for a parallel study.

Future Developments

Neuropsychological bibliography has a long tradition of segregation of MEG and EEG time series into frequency bands. In order to continue with research related to this work, the first thing that rises up is perhaps, verifying if segregation shall improve performance in classification correct ratio and most importantly to see if a later study with EnRFE technique might underline relevant SL nodes from a functional connectivity perspective.

A fast check, of how frequency band segregation might influence classification can easily be performed. Frequency band isolation has been done to original data (no out layers rejection) in order to evaluate expectable improvements. Input raw data was filtered with the following linear phase filters[16]:

- Alpha1: 8-11 Hz
- Alpha2: 11-14 Hz
- Beta1: 14-25 Hz
- Beta2: 25-35 Hz
- Gamma: 35-45 Hz

After time series being filtered, SL algorithm can show identical type of synchronization matrixes as the ones studied in this word. Next figure shows a measure of stability by means of analyzing the mean average of the Euclidean distance between all epochs in a subject. Radios variable circles denote standard deviation of previous averaged distances.

Result is absolutely decreasing with frequency bands. This result is probably due to the dynamical nature of the brain functioning. In higher frequency bands faster processes tend to be graphed by MEG SL analysis and as a complete MEG analysis must compute 10s (typically) of a signal, more different functions can arise. Lower frequency bands tend to graph slow brain processes, and therefore have more stable or more robustness in the measures. Another possible reason for this behaviour could be explained if we consider that brain connectivity happens between different functional areas and with several frequencies. Signal analysis of filtered MEG registries may hide relevant frequency components and therefore deteriorating SNR. Or maybe the effect of transitory effects because of phase irregularities of filter transfer function.

Whether the early anatomical connectivity impairment modulates profiles of functional connectivity in MCI patients is still a matter of debate (Bajo et al, 2010). Even the full definition of what Alzheimer Disease is, is now under revision. Therefore a very interesting thing to do is to compare the evolution of the patients submitted to this test as MCI subjects and therefore check whether AD has relation with predictions stated in this work.

Regarding EnRFE results, it is interesting to show how even though while features where being eliminated 100 % correct ratio result was obtained, the relevant information is not quite in that figure, since the loss of information can for sure have drastic worsening effect on

[16] SL index cannot allow an accurate enough estimation under 8Hz, and therefore alpha1 band starts at that frequency.

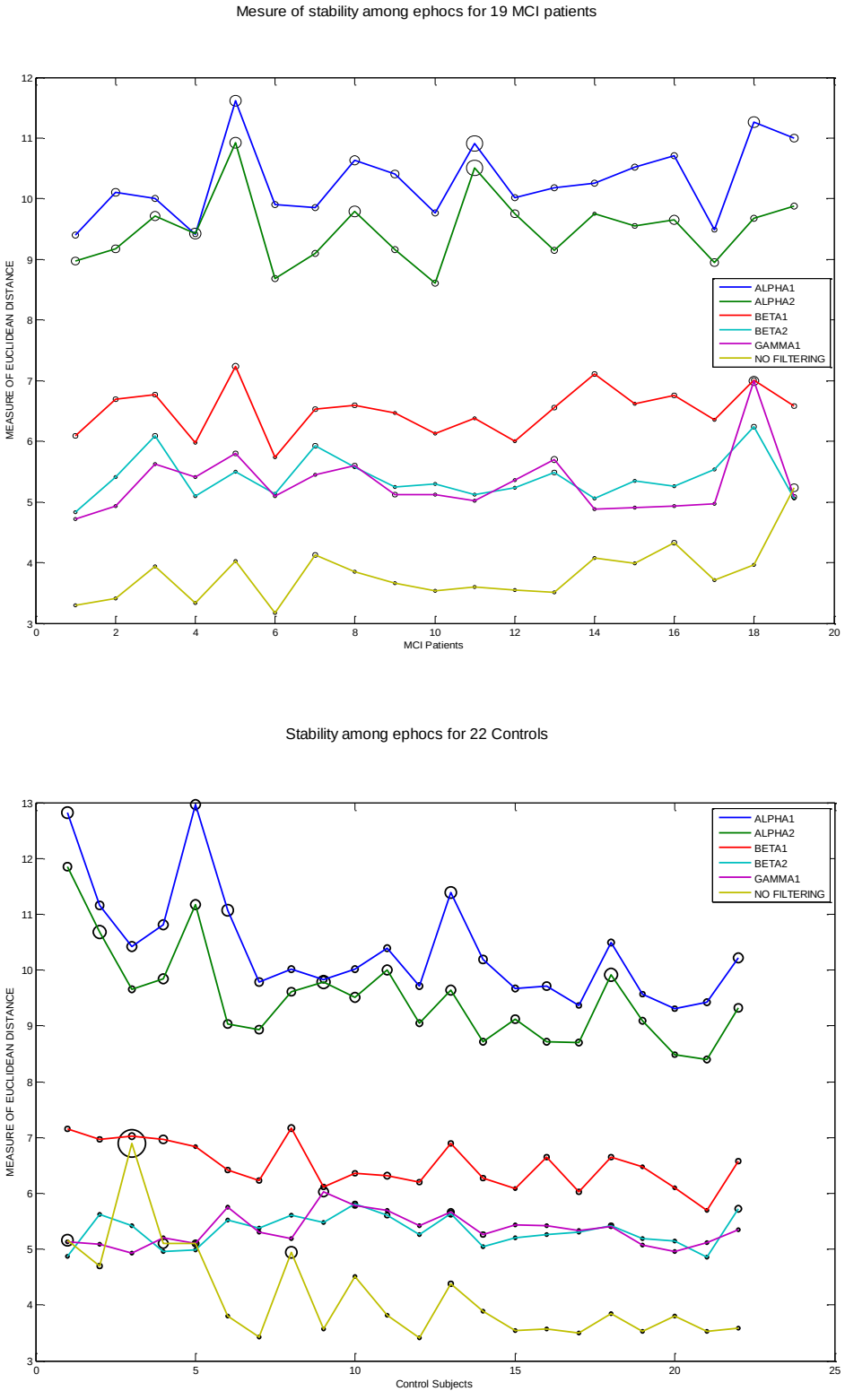


Figure 39 Measure of stability of SL values, for single subjects for different frequency bands.

generalization errors of the final SVM. However one development that could be performed is to reduce data dimension until the number of dimensions is smaller than the number of samples. In our case, as shown in table 1, we have 660 samples of control subjects, and 753 samples of MCI member subjects.

A feature reduction could be performed in order to reduce the number of features from 10878 to bellow 660. This way a widely used measure of clustering could be used to evaluate out layers before the training process starts. This measure is known as Mahalanobis distance.

Mahalanobis Distance takes into account variance covariance between the variables and hence removes problems related to scale and correlation, that are inherent between Euclidean distance.

$$D(x,y)= D(x,y) = \sqrt{(x-y)^T C^{-1}(x-y)} \quad (15)$$

With C being the covariance between the variables involved.

BIBLIOGRAPHY

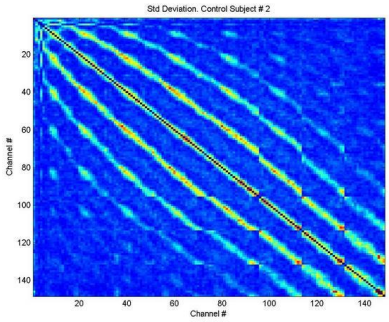
- Bajo R, Maestú F, Nevado A, Sancho M, Gutiérrez R, Campo P et al., ,2010. Functional Connectivity in Mild Cognitive Impairment Duting a Memory Task: Implications for the Disconnection Hypothesis , J Alzheimers Dis
- Braak H, Braak E, ,1991. Neuropathological stageing of Alzheimer-related changes , Acta Neuropathol 82:239-259
- Bucolo, Sapuppo Shannahoff-Khalsa,2008. From Synchronization to Network Theory: A Strategy for MEG Data Analysis. , 16th Mediterranean Conference on Control and Automation Congress Centre
- Burges, C. J., ,1998. A tutorial on support vector machines for pattern recognition , Data mining and knowledge discovery 121-167
- Chih-Wei Hsu, ,2003. A Practical Guide to Support Vector Classification , http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf
- Flicker, C., Ferris, S.H., Reisberg, B.,1991. Mild cognitive impairment in the elderly: predictors of dementia , Neurology. 1006-9
- Friston, K. J., Functional and effective connectivity in neuroimaging: a synthesis ,1994. , Human Brain Mapping 2 56-78
- Grundman M, Petersen R.C., Ferris S.H., Thomas R.G., Aisen P.S. Bennett D.A.,2004. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials , Arch Neurol 61:59-66.
- Hall, P., Marron, J. S. Neemanm, A.,2005. Geometric representation of high dimension, low sample size data , Journal R. Statistics Society B 427-444
- Jessell TM, ,2000. Principles of neural science. , McGraw-Hill, Health Professions Division 1227-1246.
- Karel Z., ,1994. Contrast limited Adaptive Histogram Equalization , Graphic Gems IV Sandigego Academic Press Professional
- Nouretdinov, I.,, et al.,,2010. Machine learning classification with confidence: Application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression, , NeuroImage doi:10.1016/j.neuroimage.2010.05.023

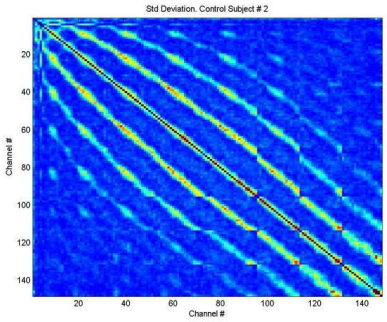
- Petersen RC, ,2004. Mild cognitive impairment as a diagnostic entity , J Intern Med 256:183-194
- Phillips P. Jonathon, ,. Support Vector Machines Applied to Face Recognition , National Institute of Standards and Technology
- Shafer G., Vovk V.,2008. A Tutorial on Conformal Prediction , Journal of Machine Learning Research 371-421
- Stam CJ, de Haan W Daffertshofer A, Jones BF, Manshanden I, van Cappellen et al.,2002. Graph theoretical analysis of magnetoencephalographic functional conectivity in Alzheimer's disease , Brain 132:213-224
- Vapnik, Vladimir N., ,1998. Statistical Learning Theory , Wiley, New York
- Xue-wen Chen, J. Cheol Jeong, ,2007. Enhanced Recursive Feature Elimination , 6th International Conference on Machine Learning and Applications . DOI 10.1109/ICMLA.2007.35

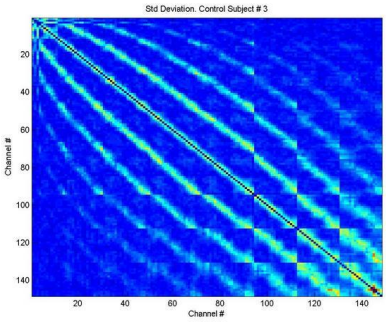
APPENDIX A: SYNCHRONIZATION LIKELIHOOD VALUES

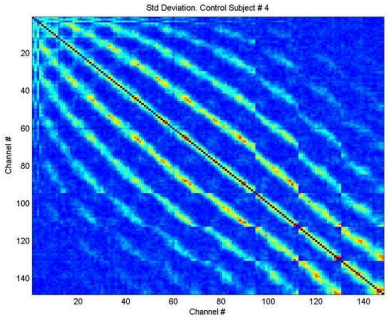
Mean reduction seems reasonable as a dimension reduction method, but in search of a way to qualify data we could compute the standard deviation of each SL value among all epochs and build this way a single image which could show how stable every SL value is during al 35 epochs.

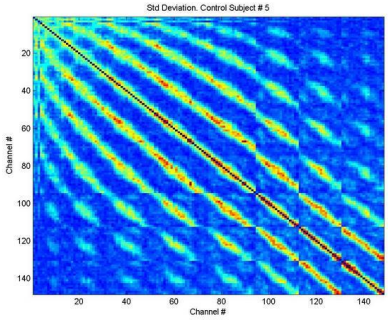
This is a valid approach to asses with the stability of each subject. Cut unfortunately it wont get the process any further than the less blue the less badly sort of qualitative approach.

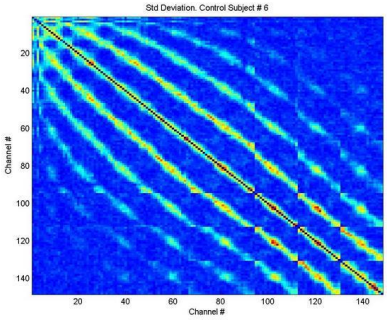


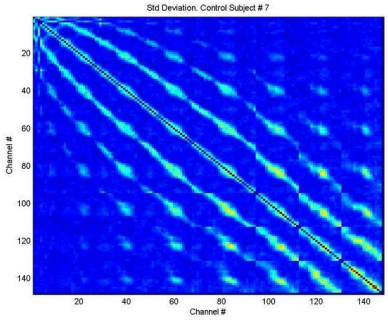


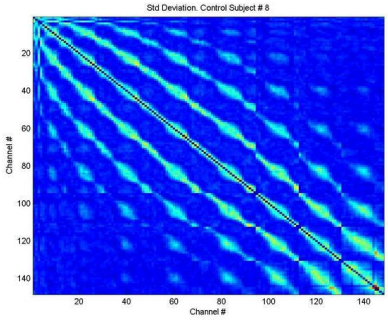


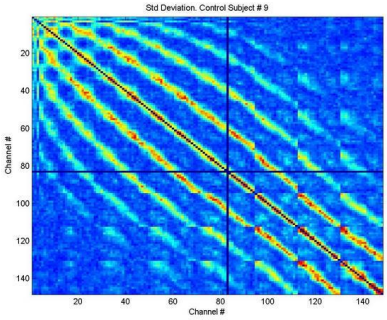


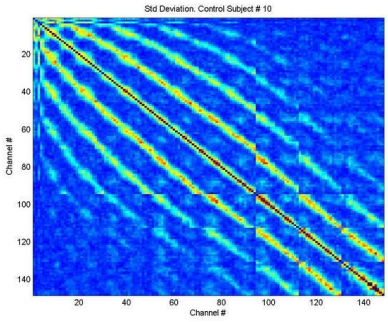


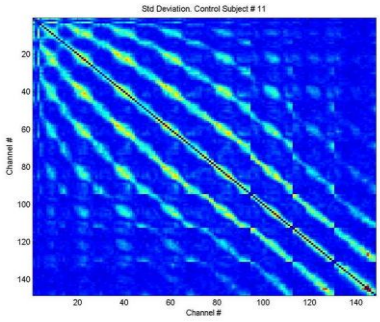


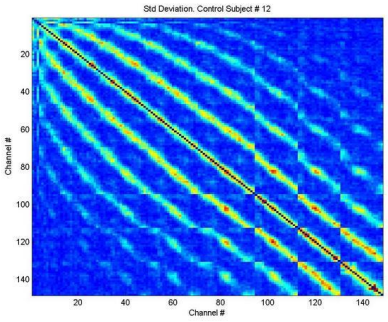


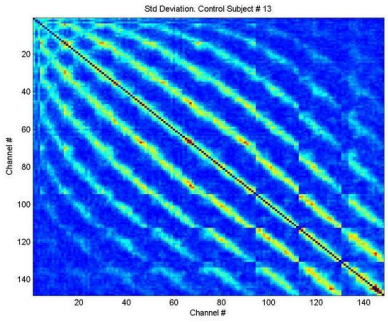


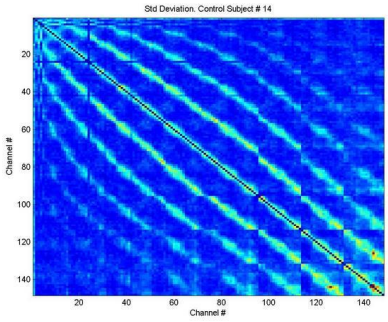


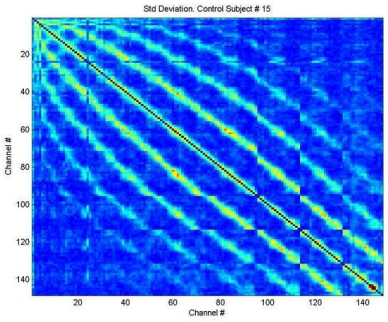

Std Deviation. Control Subject # 10
Channel #
Channel #

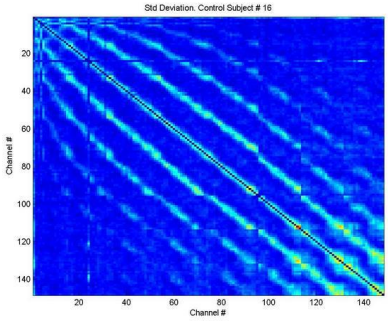

Std Deviation. Control Subject # 11
Channel #
Channel #

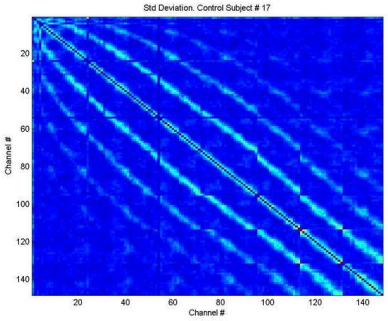

Std Deviation. Control Subject # 12
Channel #
Channel #

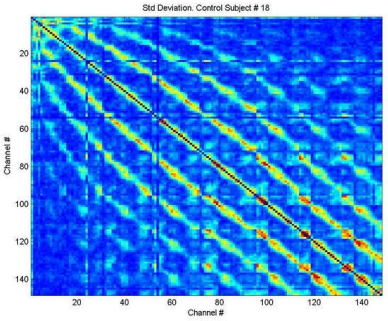

Std Deviation. Control Subject # 13
Channel #
Channel #

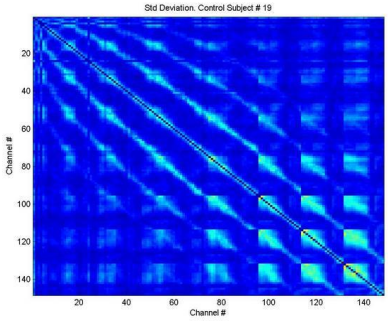

Std Deviation. Control Subject # 14
Channel #
Channel #

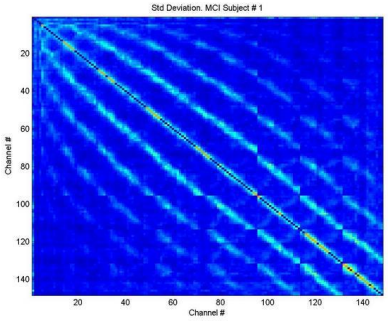

Std Deviation. Control Subject # 15
Channel #
Channel #

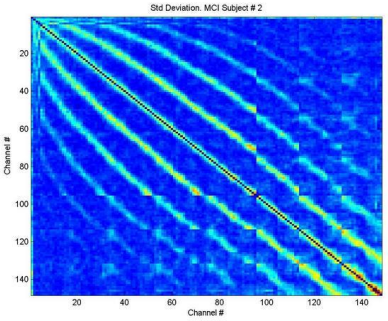

Std Deviation. Control Subject # 16
Channel #
Channel #


Std Deviation. Control Subject # 17
Channel #
Channel #

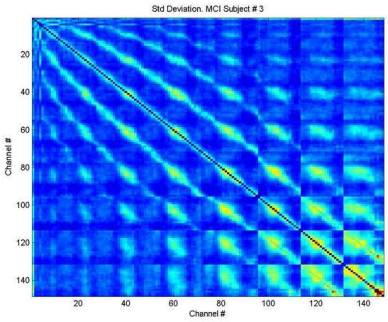

Std Deviation. Control Subject # 18
Channel #
Channel #

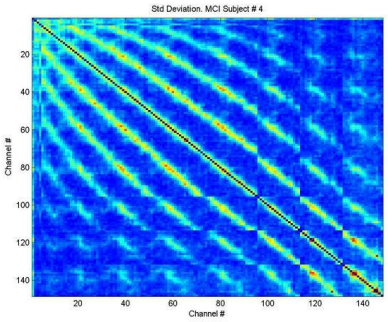

Std Deviation. Control Subject # 19
Channel #
Channel #

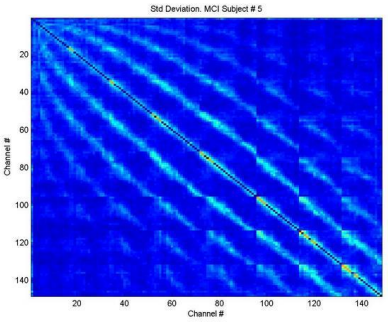

Std Deviation. MCI Subject # 1
Channel #
Channel #

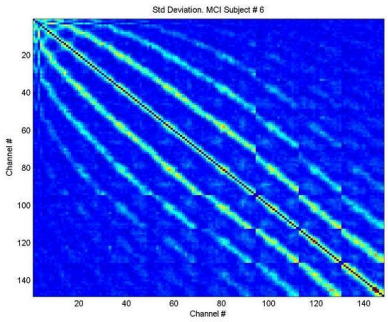

Std Deviation. MCI Subject # 2
Channel #
Channel #

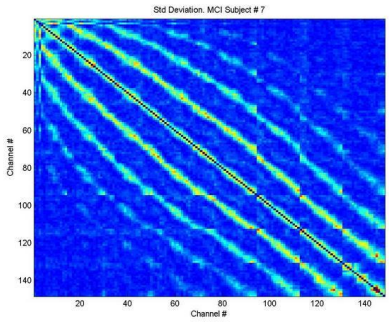
Appendix A: Synchronization Likelihood Values

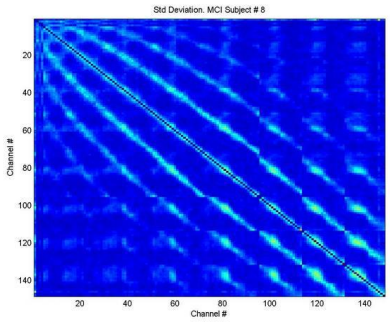


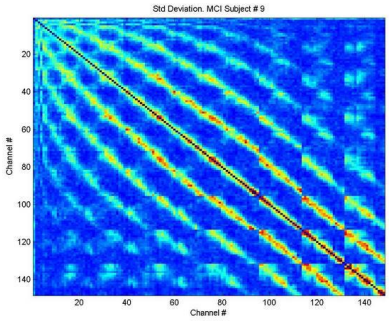


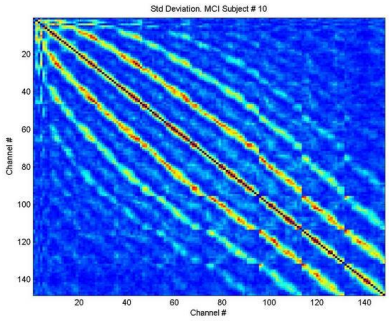


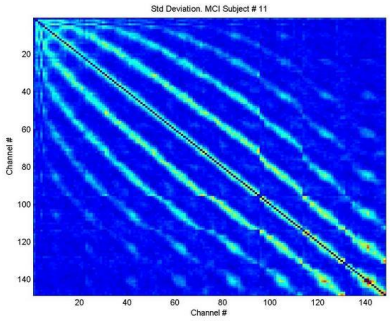


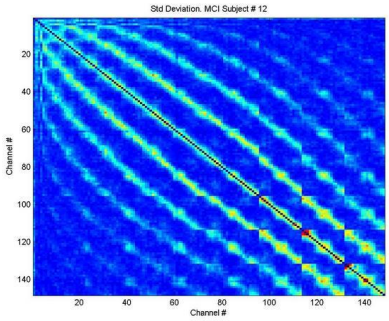


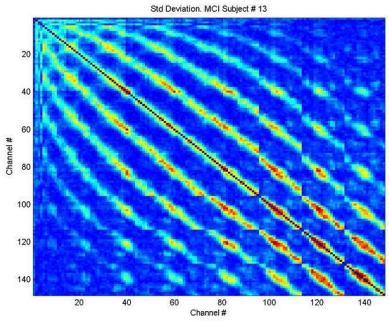


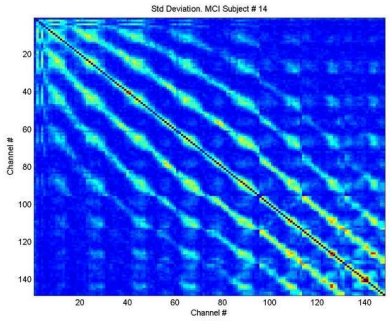


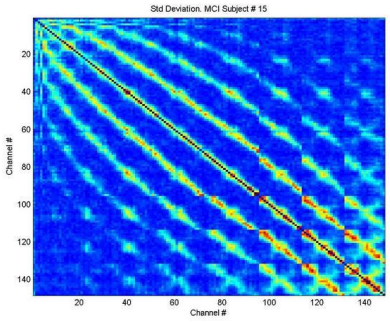


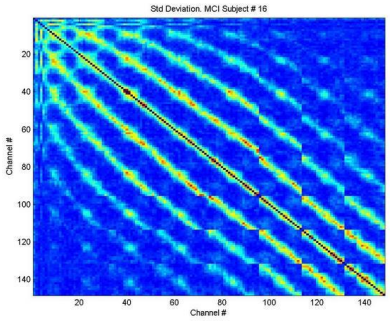


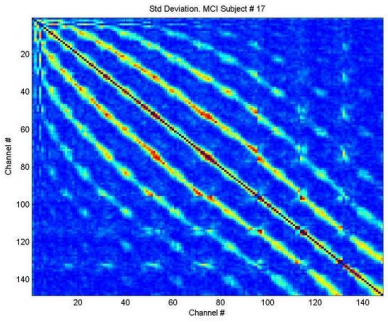


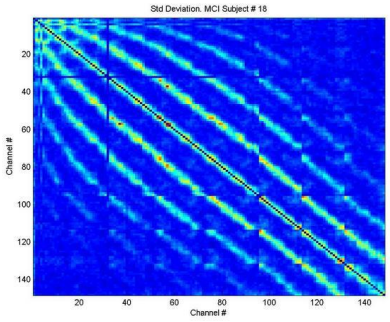


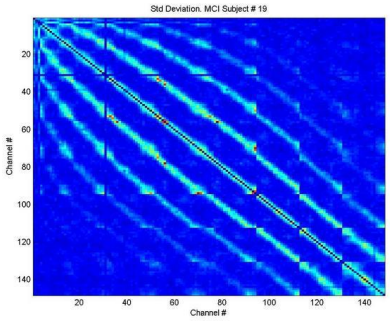


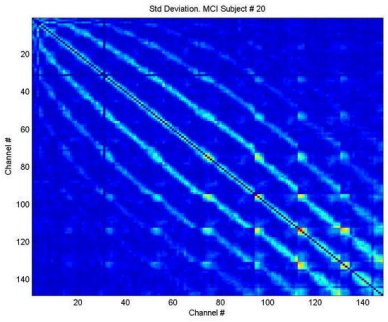

Std Deviation. MCI Subject # 15
Channel #
Channel #

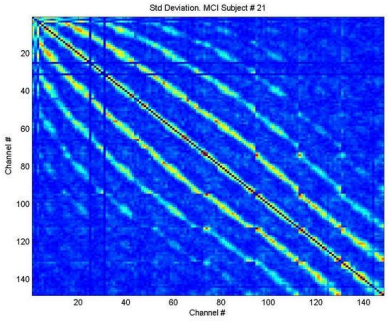

Std Deviation. MCI Subject # 16
Channel #
Channel #

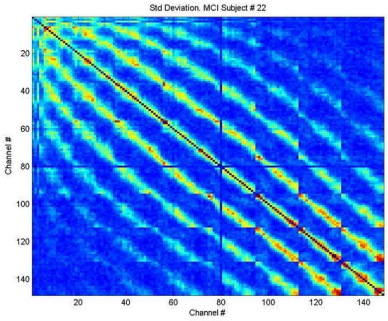

Std Deviation. MCI Subject # 17
Channel #
Channel #


Std Deviation. MCI Subject # 18
Channel #
Channel #

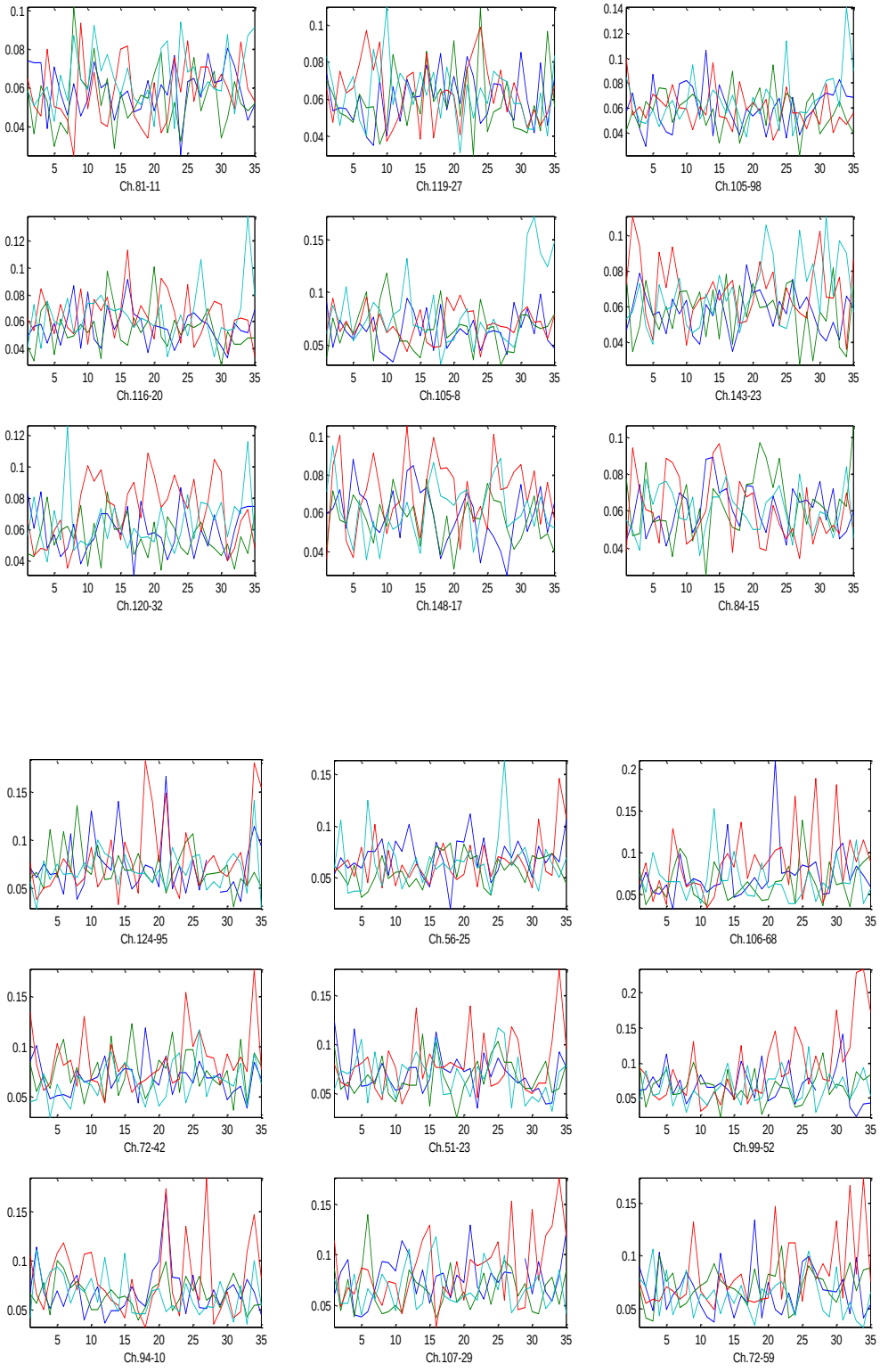

Std Deviation. MCI Subject # 19
Channel #
Channel #


Std Deviation. MCI Subject # 20
Channel #
Channel #

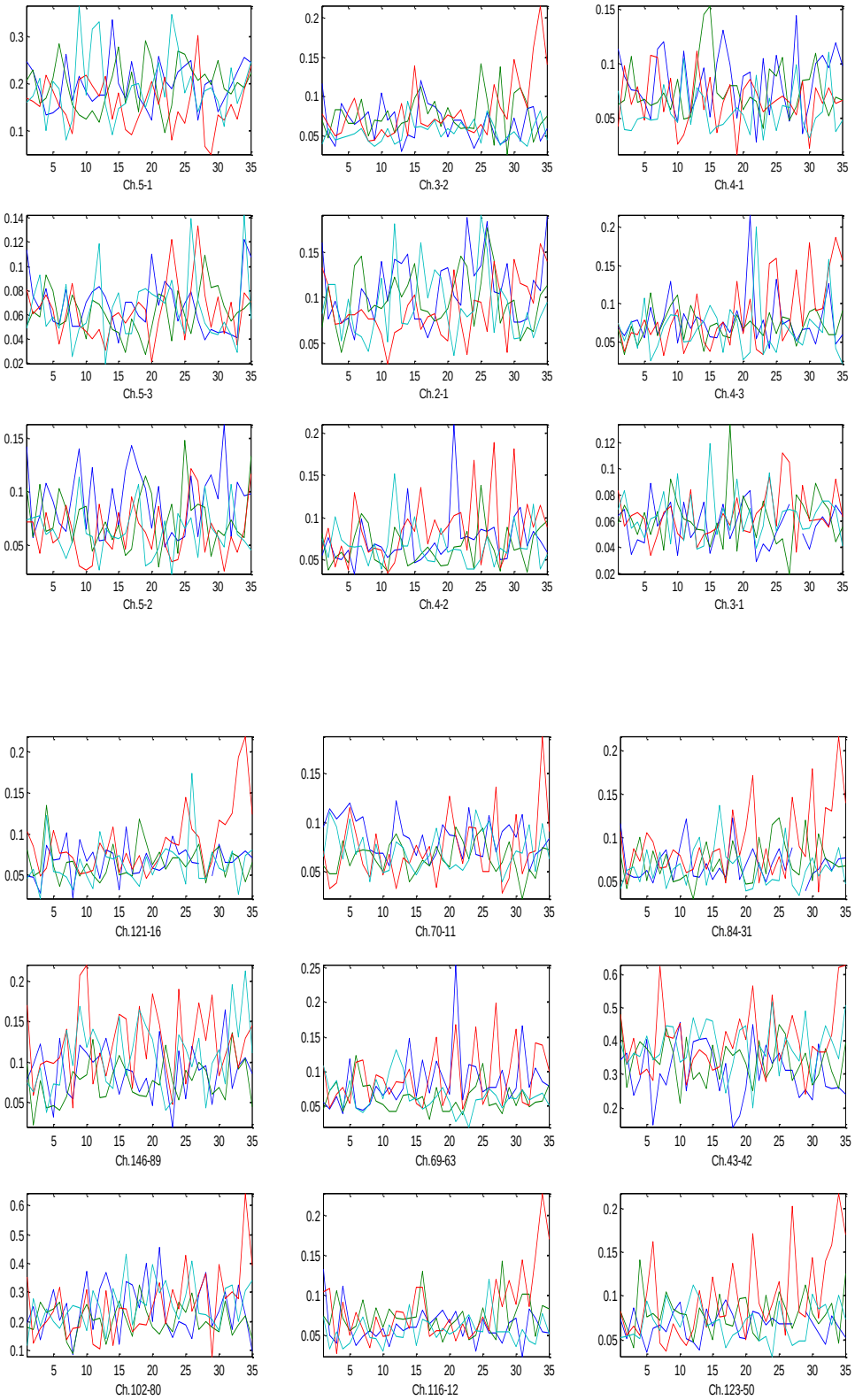

Std Deviation. MCI Subject # 21
Channel #
Channel #


Std Deviation. MCI Subject # 22
Channel #
Channel #

Different Synchronization Channel Analysis over 4 control subjects



Different Synchronization Channel Analysis over 4 MCI subjects



APPENDIX B: ABOUT MEG

Magnetoencephalography

Starting with the first development of stable SQUID detectors in 1965 by J. E. Zimmerman, Magnetoencephalography has become one of the three main techniques used for the study of the human brain from a functional approach. These are: Functional Magnetic Resonance Imaging (fMRI); Electroencephalography (EEG), Positron Emission Tomography (PET) and MEG. Different characteristics regarding each of the previous techniques tend to determine strengths and weak points among each one. And however it is a reasonable young technique that is mainly used in neuroscience research centres worldwide, it should be noted that MEG is not the answer to all the questions in cognitive neuroscience and suffers from several limitations (see Maestú et al, 2008).

Figure 40- Supine MEG setup

The Magnetoencephalograph non-invasively measures the MEG signals produced by electrically active tissue of the brain. These signals are recorded by a computerized data acquisition system and then interpreted by trained physicians to help localize these active areas.

MEG allows simultaneous measurement of 148 (306 with later developments) MEG signals inside a cryogenic Dewar vessel. The gantry, which supports the Dewar, the patient chair is operated inside a magnetically shielded room (MSR). MEG electronics unit outside the magnetically shielded room reads out the sensor outputs through the filter unit, digitizes the signals and controls the operation of the sensors.

A head Positioning System Indicator (PSI) system and a three-dimensional digitizer are also included in the system to determine the position of the head with respect to the sensor array.

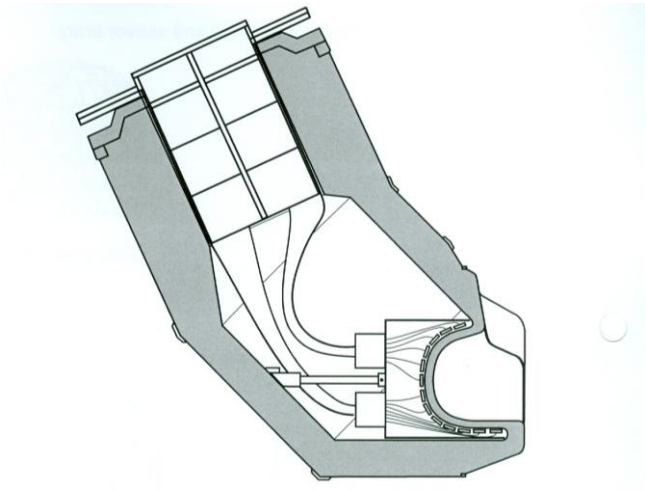


Figure 41. Cross section of the Dewar.

The probe unit construction is shown schematically in figure 41. The 148 sensors comprise magnetometers that measure de B_z component which is perpendicular to the surface of the detector of the field. The helmet-shaped cryogenic Dewar is a vacuum-insulated vessel to keep the liquid Helium necessary for cooling the SQUID sensors to 4.2 K. It is a double wall structure with vacuum gap and additional thermal radiation shielding in between. The neck plug of the probe unit also provides thermal insulation.

Gantry, bed and chair comprise a system to position the head of the subject/patient in the sensor array inside the Dewar. The chair provides seated measurement position which is the de-facto standard in cognitive studies.

MEG highlights in its context

When working with an EEG setup, neural based electrical currents are attenuated and distorted when passing through biological tissues, while reaching electrodes. Conversely, due to the fact that the magnetic field is not attenuated or distorted by biological tissues it is possible to create brain activity models at a source space.

The homogeneity of the magnetic field, in comparison with the inhomogeneity of the distribution of the electrical currents at the scalp (i.e. differences in skull thickness) may mean that parietal electrodes have higher gain than frontal electrodes in EEG. In contrast skull thickness does not affect the relation between source strength and sensor signal MEG.

No need of reference in MEG studies which facilitate synchronization or coherency analysis. Connectivity measures and source reconstruction solutions depend to some extent on the positioning of the reference channels.

Regarding the power of the higher frequency bands, the resistance of the biological tissues to electrical current produces severe attenuation of the EEG signal in the case of the high-gamma band. MEG, on the other hand, is much more sensitive to this range of frequencies.

Moreover, in the last decade there is an increasing evidence of the close relation between the gamma band and cognition.

Attending non invasiveness: with MEG records, the magnetic field induced by neuronal currents without the need to inject any substance or to expose the brain to high magnetic fields. There is no limit to the number of scans that can be performed on a single individual with MEG as opposed to fMRI and PET. This is particularly critical with children.

MEG is the only technique able to accurately combine both the spatial and temporal dimensions when measuring brain activity. fMRI and PET have better spatial resolution; however, their temporal resolution is rather poor in comparison to MEG which shows a high temporal resolution (as high as EEG) and spatial resolution validated against the Wada test and against electrocorticalstimulation (Maestú et al, 2002; Maestú et al., 2004c; Papanicolaou et al, 1999; Papanicolaou et al., 2004). High spatio-temporal resolution is particularly important in the analysis of brain connectivity in source space.

fMRI and PET measure neuronal activity indirectly. When a group of neurons become activated there is a local increase (about 4000ms after activation) in blood flow. With such a temporal resolution it is not possible to measure oscillatory activity in the most relevant frequency bands. The analysis of oscillatory activity afforded by MEG allows for estimation of phase synchronization indices between brain regions.

MEG scanners only cover the head of the subject leaving the face and body free. No sound or movement is produced by the equipment. Therefore, MEG scanning is a significantly more comfortable experience, than other techniques, making it suitable for recording responses to sound or mechanical stimuli, and also ensuring a much easier methodology with children or dementia patients.

Recently, the degree of synchronization of brain signals recorded with MEG from patients with MCI against that of healthy controls during a memory task. Synchronization Likelihood, an index based on the theory of nonlinear dynamical systems, was used to measure functional connectivity during the memory task patients show higher inter-hemispheric synchronization than healthy controls between left and right anterior temporo-frontal regions (in all studied frequency bands) and in posterior regions in the band. On the other hand, the connectivity pattern from healthy controls indicated two clusters of higher synchronization, one among left temporal sensors and another one among central channels. Both of them were found in all frequency bands. In the band, controls showed higher SL values than MCI patients between central-posterior and frontal-posterior channels and a high synchronization in posterior regions. The inter-hemispheric increased synchronization of values could reflect a compensatory mechanism for the lack of efficiency of the memory networks in MCI patients. Therefore, these connectivity profiles support only partially the idea of MCI as a disconnection syndrome, as patients showed increased long distance inter-´hemispheric connections but decrease in antero-posterior functional connectivity.

In the study of such interactions between brain areas the concept of functional connectivity has emerges referring to the statistical interdependencies between physiological time series recorded in various brain areas simultaneously (Aertsen et al., 1989). Several statistical

techniques to study interactions have been developed both in time and frequency domains, in both linear and non linear frameworks (for a extensive review see (Pereda 2005).

SQUID Electronics

The Superconducting Quantum Interference Device (SQUID) is the only sensor with sufficient sensitivity for biomagnetic measurements. The SQUID is a transducer that converts neural magnetic flux into electric signals. The electronics boards include preamplifiers for SQUID readout inside the probe unit, and main boards usually contain analog to digital converters (A/D), digital to analog converters(D/A) and a digital signal processor for feedback loop, as well as adjustable digital anti-aliasing low-pass and high-pass filters. The main boards reside inside the main electronics cabinet and are connected to the real-time data acquisition computers for control and for data forwarding.

Radio frequency interference shielding of the SQUID electronics is provided using the filter unit which is an appropriate cabinet outside the magnetically shielded room with feed through filters for all cables and isolation of power lines. The signal cables between the preamplifier boards on the top plate of the Dewar and the filter

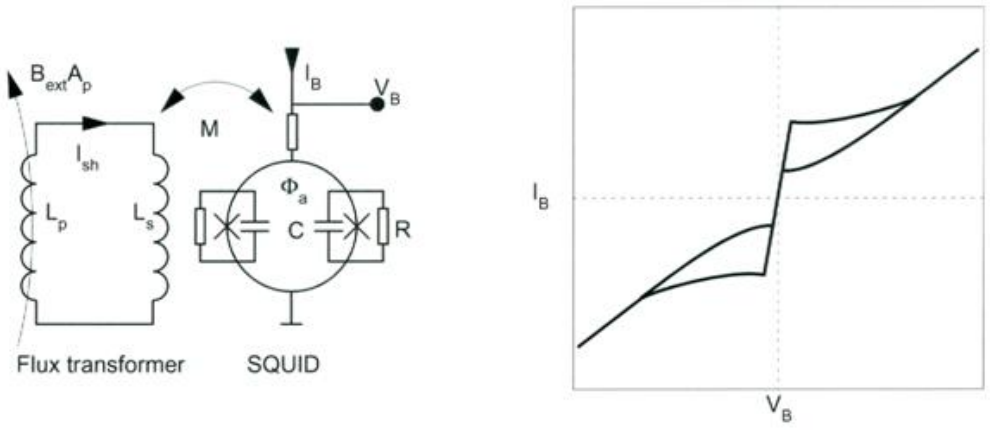


Figure 42 Left: schematic of a SQUID sensor. Right: Current-voltage characteristics of a typical SQUID sensor

The external magnetic field is not sensed directly by the SQUID; rather, it is coupled to the SQUID detector by means of a flux transformer. The flux transformer consists of two coils: a pickup coil that gathers the flux, and a signal coil that couples it into the SQUID. This has the advantage of increasing the field sensitivity by increasing the effective sensor area.

The SQUID is formed out of a superconducting ring, interrupted by two weak links or so called Josephson junctions. Without these interruptions the external magnetic fields, such as those generated by the brain would have no detectable effect on the superconducting ring. These links consist of a microscopical isolating layer that is thin enough to ensure that the ring will maintain its superconducting properties, but only up to a certain limit.

When operating the SQUID, a small current (bias current) is fed through the SQUID ring and the voltage over the device is measured. When the current through the SQUID is small, no voltage appears because the ring is totally superconducting. Above a critical value (critical current) a voltage drop appears. The apparent critical current of the SQUID depends on the

magnetic flux threading the ring. Thus, maintaining the bias current at a suitable level, a small change in the magnetic flux coupled from the external source via the flux transformer will change the point where the ring loses its superconductivity and voltage drop appears. This results in a modulation of the voltage as a function of magnetic field.

The characteristics consist actually of a family of curves for each value of magnetic flux threading the ring. However, the dependence on the magnetic flux is periodic and it is customary to plot only the extreme curves. As the magnetic flux through the ring changes the shape of the current vs. Voltage curve changes continuously between the two extrema. The period with which the behaviour repeats itself is one flux quantum...

Near the origin, the SQUID is in superconducting state and current can floss through without a voltage loss. In this case there are a deliberately added small series resistance which causes a finite slope near the origin as seen in the figure.

Data Acquisition System

The data acquisition system includes interface units to import and export digital signals. It imports the 148 MEG channels and in addition it handles the control of the MEG electronics. The data acquisition system consists of parallel real-time computers that are connected to a single acquisition workstation.

It may be interesting to mention a couple of words about the electromagnetic compatibility of the systems. Parts of the Elekta system must be permanently installed inside a magnetically shielded room. Prior to installation, a magnetic site survey and determination of necessary magnetic shielding must be performed for each installation site as a normal part of the site planning. The magnetic shield also

Shared Sources

As it has been explained already, a group of approximately 10^5 neurons are to be considered the source for the neuromagnetic field captured with MEG sensors. The fact that a large number of sensors will improve signal quality will push higher the number of sensors in the setup, nevertheless as some sensors will unavoidably close to other, sampled signal will suffer from shared source problem. Actually, this problem represents a field for research by its own.

The particular way in which sensors are numbered in the MEG setup, this is, following a spiral, will generate parallel to the main diagonal stripes on the sync matrix. According to channel numbering close sensors will typically have higher SL values. This effect will be considered a spurious component of the signal, and it will be part of the job of the classifier to avoid misinterpreting this effect.

Next figure shows how spatial distribution of sensors close to the patient’s head produces stripes or wave effect parallel to the main diagonal.

Typical sensor configuration

Figure 43 Typical sensor spatial numbering configuration